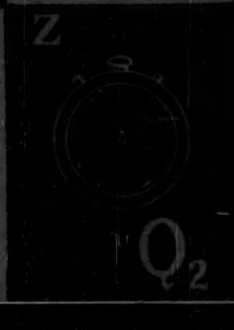
RESEARCH QUARTERLY

OCIA[illegible] [illegible]LTH
[illegible]CATION [illegible]ATION

Vol. 30, No. 2
May 1959



THESIS ABSTRACTS IN HEALTH, PHYSICAL EDUCATION, AND RECREATION, 1956 and 1957

This continuation of the 1954-55 compilation with the same title was prepared by the Thesis Abstracts Committee of the Research Council of the AAHPER as a service to those desiring up-to-date information concerning recent research developments in these areas. The 1956-57 edition has a complete author and subject index. The volume is available, in mimeographed form, for $2.00 postpaid. Copies of the *Thesis Abstracts in Health, Physical Education, and Recreation* for 1954-55 are also available for $2.00 per copy.

Send orders to: Alfred W. Hubbard
300 Huff Gym
University of Illinois
Urbana, Illinois

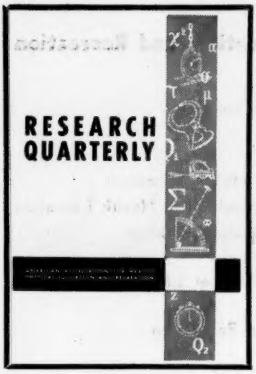
Make checks payable to:
University of Illinois
12-36-10-170

Third Annual Bibliography of Completed Research in Health, Physical Education, and Recreation

Covering research reported between September 1, 1956, and August 31, 1957, this compilation is available for $1.00.

Send orders to: AAHPER
1201 16th Street, N.W.
Washington 6, D. C.

Both of these series are being discontinued. The thesis materials, together with journal references, are being consolidated in a new, annual publication. AAHPER will print and distribute these valuable reference materials under the title Completed Research in Health, Physical Education, and Recreation. *Each of these annual publications will record research production for the year immediately preceding and will be completely and conveniently indexed. Watch for an announcement of the first issue covering 1958.*



The Research Quarterly
of the American Association for Health, Physical Education, and Recreation
A department of the National Education Association
1201 Sixteenth St., N.W., Washington 6, D. C.

VOL. 30, NO. 2 **MAY 1959**

CONTENTS

Published in March, May, October, and December by the American Association for Health, Physical Education, and Recreation (a department of the National Education Association), 1201 Sixteenth Street, N.W., Washington 6, D. C. Subscription available to libraries, institutions, and departments at $5.00 per year. *Subscriptions to individuals are available only through professional membership in AAHPER.* Of the Professional Membership ($15.00 per year), $4.00 covers subscription to the *Quarterly.* Of the Student Professional Membership ($6.00) per year), $1.50 is for subscription to the *Quarterly.* Single copies $1.25. Send membership dues or subscription fees to AAHPER Circulation Department, 1201 Sixteenth Street, N.W., Washington 6, D. C.

Manuscripts are to be sent to the *Research Quarterly,* AAHPER, 1201 Sixteenth Street, N.W., Washington 6, D. C. Authors should follow the form prescribed in the "Guide to Authors" that appears each year in the October issue of the *Quarterly.*

Second class postage paid at Washington, D. C., and at additional mailing offices.

An Experimental Study of the Effect of Water Ingestion upon Athletic Performance

LANE BURTON BLANK
University of California
Los Angeles

Abstract

Thirty-three collegiate track and field men, divided into three random groups, ran 27 individual 220-yard time trials under three alternating conditions: (1) no liquid consumption within the hour preceding the trial; (2) consumption of one pint of water five minutes prior to the trial; and (3) free consumption of water five minutes before the recorded run. Differences in performance under the three conditions were discovered to be insignificant statistically as well as inconsistent in their direction. Thus the null hypothesis concerning the effect of water ingestion upon performance was supported.

MOST ATHLETIC COACHES follow the practice of advising their charges sternly not to consume liquids prior to competing in a game, match, or event. This is usually done on the theory that such habits may inhibit performance and/or leave some residual degree of infirmity, at least on a temporary basis.

Bresnahan, Tuttle, and Cretzmeyer (2) are typical of the track and field authors warning against the drinking of water just prior to competition:

> Some coaches and trainers tend to withhold water from competitors for a considerable time before competition. It is agreed that the drinking of water immediately before a contest might prove troublesome. However, ample water can safely be allowed up to an hour before competition time.

Many athletes, and even some coaches and researchers, however, have been inclined to question this warning on the grounds that the participant may not really be impeded by taking a preliminary drink of water or other liquid and, in fact, may even be aided in his performance by engaging in such a practice.

In the realm of related literature, the experiments of Little, Strayhorn, and Miller have provided interesting, though relatively small, background. In essence, they set out to examine the "common impression that distention of the stomach with large amounts of water decreases the capacity for strenuous exertion" (4).

The investigations of these researchers at the University of North Carolina were twofold, in the use of treadmill activity and performance in running and swimming. Quantities of water ranging from 1 to 1.5 liters were ingested by subjects during the few minutes immediately preceding strenuous trials, and the principal findings were reported as follows: (1) "water drinking

produced sensations of discomfort and some disinclination for exercise, but there was no discomfort DURING the exercise"; and (2) "in no case was performance adversely affected by the prior drinking of a large volume of water." In addition, it was reported that "unpublished data . . . indicate that a football game, even when played in moderately cool weather, may result in sufficient dehydration to affect performance adversely when water drinking is too severely restricted" (4).

It was the purpose of the current study to determine the effect, if any, of the ingestion of water by athletes immediately prior to participation in a strenuous physical activity. In undertaking the study, the writer has hypothesized that there is no significant difference between performances under the various experimental conditions. The present investigation has been delimited in that: (1) research has been confined to performance in a single track event, the 220-yard dash; (2) only members of a collegiate track and field team were used as subjects.

Definition of Terms

The so-called *control* condition of the experiment refers to a time trial in which each participant has drunk no liquid within the hour immediately preceding.

The *experimental* condition denotes a time trial preceded by the ingestion of the specified quantity of water.

The *free* condition is one in which the participants are permitted to drink any quantity of water immediately preceding the time trial and are encouraged to take some water.

Warming-up consists of the participation of an athlete in preliminary physical activity prior to engaging in competitive or practice operations (1).

Procedure

The subjects participating in the experiment were volunteers from the membership of a collegiate track and field squad, and they represented every event in the program. From the original 35 volunteers, two were dropped at random to reduce the total number to a multiple of three. Then the 33 remaining subjects were divided by random means into three groups of 11 each for the purposes of the experiment.

On the initial day of the experiment each of the three groups participated in 220-yard time trials under the separate conditions defined above. On the second and third days the conditions for the groups were systematically changed so that each subject performed under all situations consecutively. This alternating pattern was repeated in order to minimize the possible effects of conditioning. Tuesdays and Wednesdays of 17 consecutive weeks were chosen for the administration of the trials; however, the total number of test days was reduced to 27 by a week of academic holiday and several days of inclement weather. Tuesdays and Wednesdays were selected because they are generally the days of the week upon which the track and field athlete trains the hardest and also because it was felt that extra trials on these days would be least apt to hamper performance in competition on Saturday.

The purpose of the experiment was described thoroughly to the subjects, as

were the pertinent definitions and general procedure, before any activity was called for. The subjects were not, however, informed of any of their marks until after the conclusion of the study. Subjects were closely supervised on the field as to their consumption of liquids prior to the time trials. The participants were asked to refrain from drinking any liquids within the hour immediately preceding the timed run on the days on which their particular subgroup was running under the previously defined "control" condition. For the "experimental" condition the specified group each day was given one pint (16 fluid ounces) of tap water to drink five minutes before the trials, while the group participating under "free" conditions was urged to consume some water—the amount actually taken being measured, but its members were allowed to take any amount they desired or none at all. In addition, each subject was asked to give an all-out or maximum performance for each trial.

The trial runs were administered and timed by conventional methods employing three stopwatches for each runner's performance. Subjects performed their trials individually to eliminate the possible effect of competition, after the usual daily pattern of warming-up but before attempting any other portion of their workouts.

Two principal problems of technique and experimental condition were discovered by the investigator. First, it was felt that absolute supervision of liquid consumption under the control condition was somewhat difficult to accomplish, since an occasional subject failed to observe the full hour of prescribed restriction. Second, whether or not a particular subject gave an all-out sprint performance on each trial is open to serious doubt.

Results

From the trial times over the 220-yard distance, the means for each of the subjects and for the entire group were computed, as were the differences between each pair of experimental conditions—i.e., control vs. experimental, control vs. free, and experimental vs. free. *Control* means ranged between 22.09 and 30.96 seconds, *experimental* means between 21.96 and 30.91, and *free* means between 22.07 and 30.93 seconds.[1]

Differences between control and experimental mean times ranged from 0.00 to 0.36 seconds; differences between pairs of means for control and free trials ranged from 0.00 to 0.45; while the corresponding differences between experimental and free means were found to vary from 0.01 to 0.39 seconds. It is essential to note that the direction of the differences obtained was discovered to have no particular pattern, since approximately equal numbers of subjects performed their fastest mean trials under each of the three conditions of the experiment.

---

[1]Copies of tabular material concerning the performance of individuals will be furnished by the author upon request.

Standard deviations of the three distributions of times ranged from 0.13 to 0.64 seconds for the control situation, 0.20 to 0.52 for the experimental, and 0.17 to 0.71 for the free intake trials. Group means and standard deviations are found in Table 1.

TABLE 1.—MEAN TIMES AND STANDARD DEVIATIONS IN SECONDS FOR ALL 220-YARD TRIALS

Condition	Mean	Standard deviation
Control	24.33	2.05
Experimental	24.26	2.14
Free	24.28	2.06

An analysis of variance was performed for the various combinations of conditions, and in each instance the within variance was considerably larger than the between variance. F-ratios, of the magnitude of 0.001 to 0.01, indicate that no significant difference existed among trials performed under the three conditions of the experiment.

Consumption of water by subjects participating under free intake conditions was recorded and found to range from 2.8 to 19.1 ounces per trial among the experimental subjects on a mean basis. Only three individuals failed to drink water at all prior to any of their free condition time trials, and this failure to consume water occurred on one, two, and three occasions respectively. The maximum quantity of water ingested in a single instance was 27 ounces.

Summary

The principal experimental and statistical findings are summarized as follows:

1. Differences between means for each respective pair of conditions and for each subject ranged from 0.00 to 0.45 seconds.
2. The differences obtained exhibited no significant degree of consistency; in other words, for approximately equal numbers of participants each of the experimental conditions proved to produce the fastest times.

The group as a whole exhibited no significant differences in performances between pairs of experimental conditions.

Conclusion

The principal conclusion derived from this study was that, in view of the experimental conditions exercised, there appears to be no justification—at least as concerns performance—for either recommending or cautioning against the consumption of water prior to participating in a strenuous athletic event.

Recommendations

In view of the nature of the limitations of this study and of the conditions involved, it is suggested that additional research be carried on in the following respects:

1. In a study somewhat similar to the present one, additional athletic events and/or activities might be employed and the size of the sample increased.

2. Varying quantities of water or other liquids could be utilized as well as varying preparatory periods.

3. Subjects in a similar study might be urged to report any unusual discomforts experienced in the period of trials and such comments analyzed to determine any possible trends in their occurrence.

References

1. BLANK, LANE B. *An Experimental Study of the Effect of Selected Warmup Activities on Speed in Sprinting.* Unpublished Master's thesis. Los Angeles: University of California, 1951.
2. BRESNAHAN, GEORGE T.; TUTTLE, W. W.; and CRETZMEYER, FRANCIS. *Track and Field Athletics.* 4th ed. St. Louis: C. V. Mosby Co., 1956.
3. GUILFORD, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Book Co., 1942.
4. LITTLE, C. C.; STRAYHORN, H; and MILLER, A. T., JR. "Effect of Water Ingestion on Capacity for Exercise.' *Research Quarterly* **20**: 398-401; Dec. 1949.

(Submitted 8/8/58)

The Relationship Between the Physical Fitness Ratings of Aviation Cadets and Certain Early Life Experiences Pertaining to Physical Activity

CHARLES A. BUCHER
New York University
New York, New York

DOMINICK A. TADDONIO
University of Detroit
Detroit, Michigan

Abstract

The Army Air Force Physical Fitness Test and a questionnaire which sought data on certain early life experiences were administered to 1226 aviation cadets during World War II.

Findings revealed lack of appreciable differences in physical fitness in later years among the following cadets: (1) attended public, private, and parochial schools; (2) did and did not experience required physical education programs; (3) did and did not attend college; (4) lived in various geographical areas of the United States; and (5) lived in rural as opposed to urban environments.

Some substantial differences were noticed relative to elementary school activities and to varsity sports participation.

INVARIABLY, statements of objectives for physical education include the development of physical fitness. The means to the end is activity, which is fostered by teaching of skills, the development of attitudes, and the opportunity to participate. The means are provided through required physical education programs, intramurals, and interschool athletics. Further, such programs are said to engender future physical fitness because skills and attitudes "carry over."

Purpose of the Study

This study is an attempt to glean from a questionnaire administered to 1226 aviation cadets during World War II the relationship between certain early life experiences pertaining to physical activity and physical fitness in later years. The questionnaire was prepared by Capt. Herbert Crowley, director of physical training, Maxwell Field, Alabama.

Specifically, the study attempted to determine the physical fitness of cadets who: (1) attended public, private, and parochial schools, (2) *did* and *did not* experience required physical education programs in the elementary

school, (3) preferred one "favorite" activity to another in the elementary school, (4) attended public, private, and parochial high schools, (5) *did* and *did not* experience required physical education programs in high school, (6) *did* and *did not* participate in varsity sports in high school, (7) participated in the various varsity sports in high school, (8) *did* and *did not* attend college, (9) lived in the various geographical areas of the United States, and (10) lived the greater part of their childhood and boyhood in a rural environment as opposed to an urban environment.

Because data were available, the writers determined the physical fitness of cadets based on age.

Methodology

The Army Air Force Physical Fitness Test (sit-ups, pull-ups, and 300-yd. shuttle run) and a 50-item questionnaire were administered to 1226 aviation cadets, ages 18 to 27, in the initial phase of their training at Maxwell Field, Alabama, during World War II. Although the questionnaire was undoubtedly developed with the intent of conducting a study of this type, it was not specifically conceived for the study presented here. The writers, however, selected the physical fitness rating and certain responses to questionnaire items which they feel shed light on some interesting aspects of physical education.

PFR data were grouped, and means and standard deviations calculated.

No design was used to ensure selection of a "random" sample. Rather, the subjects included the entire cadet personnel of several training squadrons.

Geographical areas were based on the *Encyclopaedia Britannica* (1956) division of the United States (see Table 9).

Rural and urban areas were determined by U. S. Bureau of the Census definition of 0-2499 population for a rural area and 2500 and over population for urban areas. Determination of each cadet's environment, rural or urban, was made by use of Bureau of the Census statistics for 1940 as reported in the 1957 edition of *World Almanac.*

Other data merely involved tabulation of simple responses to questionnaire items.

TABLE 1.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO ATTENDED PUBLIC, PRIVATE, AND PAROCHIAL ELEMENTARY SCHOOLS

Type of School	N	Range	Mean PFR	σ
Public	1,012	34–93	62.17	11.62
Private	35	50–81	63.68	8.31
Parochial	148	38–93	61.95	10.36

TABLE 2.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO DID AND DID NOT EXPERIENCE REQUIRED PHYSICAL EDUCATION PROGRAMS IN THE ELEMENTARY SCHOOL

P. E. Required?	N	Range	Mean PFR	σ
Yes (Public, Private, Parochial)	665	34–94	62.48	10.43
No (Public, Private, Parochial)	539	34–93	62.14	10.99
Yes (Public)	608	34–93	62.23	10.50
No (Public)	404	34–92	62.25	11.13
Yes (Private)	28	51–81	64.50	7.95
No (Private)	7	52–73	61.00	8.07
Yes (Parochial)	25	48–78	63.96	9.36
No (Parochial)	123	38–93	61.55	10.57

TABLE 3.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO FAVORED CERTAIN PHYSICAL ACTIVITIES IN THE ELEMENTARY SCHOOL

Favored Activity	N	Range	Means	σ
Baseball	393	34–93	61.51	9.73
Basketball	122	38–94	62.57	9.95
Football	180	40–94	65.28	10.36
Games of Low Organization[a]	79	42–92	63.77	9.90
Handball	10	45–68	55.30	7.39
Running and Cross Country	23	47–78	62.83	9.64
Soccer	35	36–85	61.00	10.50
Softball	129	34–93	59.05	10.36
Swimming	45	44–90	62.44	12.00
Tennis	8	48–81	67.13	9.77
Track	27	48–92	66.63	11.05
Volleyball	11	47–85	62.55	11.02

[a]Includes activities such as punchball, dodgeball, stickball, etc.

TABLE 4.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO ATTENDED PUBLIC, PRIVATE, AND PAROCHIAL HIGH SCHOOLS

Type of School	N	Range	Means	σ
Public	1,064	34–93	62.35	10.60
Private	82	40–92	62.60	10.70
Parochial	50	38–94	61.10	11.80

TABLE 5.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO DID AND DID NOT EXPERIENCE REQUIRED PHYSICAL EDUCATION PROGRAMS IN HIGH SCHOOL

P. E. Requirement	N	Range	Means	σ
Yes (Public, Private, Parochial)	967	34–94	62.25	10.45
No (Public, Private, Parochial)	229	40–92	62.60	11.10
Yes (Public)	872	34–93	62.19	10.40
No (Public)	192	40–92	63.01	11.05
Yes (Private)	67	45–93	62.96	9.95
No (Private)	15	40–85	60.33	12.48
Yes (Parochial)	28	38–94	61.50	12.98
No (Parochial)	22	38–75	60.68	9.86

TABLE 6.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADET WHO DID AND DID NOT PARTICIPATE IN VARSITY SPORTS IN HIGH SCHOOL

Participation in Varsity Sports	N	Range	Means	σ
Did Participate (Letter-winners)	465	36–94	65.70	10.15
Did Participate (Non-letter winners)	271	38–93	62.00	10.40
Did not Participate	473	34–90	59.20	10.05

Note: 105 of the 1226 cadets in the study (approximately 9%) earned a varsity letter in college. The mean PFR for the group was 68.40, range 50-94, 9.35.

TABLE 7.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO PARTICIPATED IN THE VARIOUS VARSITY SPORTS IN HIGH SCHOOL

Varsity Sport	N	Range	Means	σ
Football	308	40–94	65.75	10.75
Basketball	277	36–94	64.90	9.60
Track	173	38–94	65.80	10.90
Baseball	162	36–93	63.10	10.25
Swimming	60	50–90	66.38	9.36
Tennis	60	44–81	63.90	8.88
Soccer	31	46–85	61.10	9.64
Hockey	25	42–93	63.60	11.70
Wrestling	24	54–93	69.29	10.99
Golf	20	45–75	59.50	8.06
Softball	20	38–85	60.45	11.45
Cross Country	12	42–85	68.00	11.03
Gymnastics	12	46–72	62.17	8.04
Volleyball	9	39–73	53.67	10.40
Boxing	7	57–71	65.00	4.34
Fencing	7	52–72	61.71	9.10
LaCrosse	5	55–68	62.60	4.23
Bowling	5	38–61	50.80	8.52
Handball	4	48–61	55.25	5.45

Some cadets participated in more than one sport. In such instances, the PFR was included in the grouping for each sport.

TABLE 8.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO DID AND DID NOT ATTEND COLLEGE

College Attendance	N	Range	Means	σ
Did Attend	727	35–94	62.50	10.05
Did not Attend	485	34–93	62.10	11.80

TABLE 9.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS FROM VARIOUS GEOGRAPHICAL REGIONS OF THE UNITED STATES

Region	N	Range	Means	σ
New England	295	38–93	62.60	10.57
Middle Atlantic	448	34–94	61.46	9.87
East North Central	338	34–93	61.83	9.94
West North Central	34	44–90	64.10	12.90
South Atlantic	149	35–93	63.35	10.50
East South Central	66	45–93	64.75	11.15
West South Central	24	48-81	62.29	9.67
Mountain	10	40–78	65.00	11.81
Pacific	12	50–81	66.00	9.63

New England: Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, Connecticut.
Middle Atlantic: New York, New Jersey, Pennsylvania.
East North Central: Ohio, Indiana, Illinois, Michigan, Wisconsin.
West North Central: Minnesota, Iowa, Missouri, North Dakota, South Dakota, Nebraska, Kansas.
South Atlantic: Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida.
East South Central: Kentucky, Tennessee, Alabama, Mississippi.
West South Central: Arkansas, Louisiana, Oklahoma, Texas.
Mountain: Montana, Idaho, Wyoming, Colorado, New Mexico, Arizona, Utah, Nevada.
Pacific: Washington, Oregon, California.

TABLE 10.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS WHO LIVED MOST OF THEIR CHILDHOOD IN RURAL AND URBAN ENVIRONMENTS

Environment	N	Range	Means	σ
Rural (0– 2,499)	172	34–93	62.82	11.06
Urban (2,500– 9,999)	176	40–89	62.82	9.17
Urban (10,000– 24,999)	154	34–93	64.01	11.13
Urban (25,000– 49,999)	106	40–89	63.38	8.96
Urban (50,000– 99,999)	99	40–93	61.91	10.57
Urban (100,000–249,999)	102	40–89	61.96	11.41
Urban (250,000–499.999)	74	34–89	61.00	10.64
Urban (500,000–999,999)	99	34–94	60.51	11.41
Urban (Over 1,000,000)	205	34–82	61.07	10.43
(Composite over 2,500)	847	34–94	62.47	11.55

TABLE 11.—MEAN AAF PHYSICAL FITNESS RATINGS OF CADETS ACCORDING TO AGE

Age	N	Range	Means	σ
18	63	35–90	62.50	13.00
19	256	38–93	63.10	9.80
20	322	34–94	63.17	9.94
21	250	34–93	62.05	10.43
22	113	36–93	61.42	10.71
23	65	36–93	63.30	10.45
24	46	40–90	62.00	12.20
25	28	41–90	59.75	12.32
26	50	40–92	57.70	11.85
27	20	40–81	62.70	12.35

Discussion

The data presented should be interpreted with two thoughts in mind. First, the sample is not random by design, and therefore sampling statistics in terms of "statistically significant differences" have not been applied. Second, where differences exist in mean PFR's, consideration should be given to their "practical significance." For example, how much "practical significance" should be attached to a difference in mean PFR's of 2 points—a difference represented by approximately 2 sit-ups, one-half pull-up, and one-half second in the 300-yd. shuttle run? (See Appendix for AAF PFR Scoring Table.)

The lack of appreciable differences in mean Physical Fitness Rating among cadets who attended public, private, and parochial schools (Tables 1 and 4) seems to indicate that there are no factors in the curricular or extracurricular activities of these types of schools which make for superiority in physical fitness in later years (through age 27). Likewise, the lack of appreciable differences in mean PFR between cadets who had and those who had not experienced required physical education programs (Tables 2 and 5) seems to indicate that a "requirement" does not ensure superiority in physical fitness in later years. It should be noted for interpretation, also, that the investigators made no effort to discern the quality of required programs.

Data referring to favorite activities in the elementary school (Table 3) and participation in varsity sports (Tables 6 and 7) reveal some PFR differences which may be regarded as substantial. However, regardless of the significance attached to such differences, it cannot necessarily be assumed that participation caused the differences, since these may be due to selective factors.

Physical fitness in later years does not seem to be influenced by whether or not a cadet attended college (Table 8).

With regard to geographical areas and size of communities in which cadets were reared (Tables 9 and 10), there does not seem to be appreciable differences in mean PFR which would indicate marked superiority in physical fitness in later years, despite the relatively high PFR of cadets from the

mountain states (N = 10) and the Pacific states (N = 12), since the low number of cases attaches less weight to these data. If positive factors exist in the various environments, apparently they are negated by other factors which do not differentiate outcomes in terms of physical fitness as measured by the AAF PFR.

There appears to be no consistent pattern of relationship between the mean PFR of cadets and their age (Table 11).

References

1. LARSON, LEONARD A. "Some Findings Resulting from the Army Air Forces Physical Training Program." *Research Quarterly* **17**: 161-64; May 1946.

Appendix

AAF PFR ACHIEVEMENT SCALES[a]
Achievement Scale

Sit-Ups		Pull-Ups		Shuttle-Run		Sum of Scores		P.F.R.
No.	Score	No.	Score	Sec.	Score			
114	100	24	100	41	100	300	EXCELLENT	100
108	98	23	98			294		98
102	96	22	96	42	96	288		96
96	95	21	95			285		95
90	93	20	93	43	93	279		93
85	90	19	90			270		90
81	85	18	85	44	85	255		85
77	81	17	81			243		81
73	78	16	78	45	78	234		78
69	75	15	75	46	75	225	VERY GOOD	75
66	74					222		74
64	73			47	73	219		73
62	72	14	72	48	71	216		72
60	70					210		70
58	68	13	68	49	67	204		68
56	66					198		66
54	65	12	65	50	65	195		65
52	64					192		64
50	63	11	62	51	63	189	GOOD	63
48	61					183		61
47	60			52	60	180		60
45	58	10	58			174		58
44	57			53	56	171		57
42	55					165		55
40	54	9	54			162		54
38	52			54	52	156		52
36	50	8	49			150		50
18	33			55	48	144		48
31	47					141		47

30	46			56	46	138	POOR	46
29	45	7	45			135		45
28	44			57	44	132		44
27	42	6	41	58	42	126		42
26	40			59	40	120		40
25	38	5	38	60	38	114		38
24	36			61	36	108		36
22	35	4	35			105		35
21	34			62	34	102		34
19	33	3	32	63	32	99	VERY POOR	33
17	30			64	30	90		30
15	27	2	26	65	27	81		27
12	23			66	23	69		23
9	20			67	20	60		20
6	17	1	17	68	17	51		17
3	15			69	15	45		15
1	10			70	10	30		10

[a] From *Physical Fitness Handbook,* prepared for Army Air Force distribution by Training Aids Division, Office of Assistant Chief of Air Staff, Training Headquarters, Army Air Forces, New York 16, N. Y.

(Submitted 8/24/58)

The Selection, Development, and Evaluation of Tobacco Smoking Concepts

ALEXANDER F. GALARNEAUX
Dorchester High School
Boston, Massachusetts

CLEM W. THOMPSON
Boston University
Boston, Massachusetts

Abstract

The purposes of this study were to determine the validity and accuracy of tobacco smoking concepts and their value to general education. Concepts were divided into physiological, psychological, and socio-economic categories. Separate juries of experts in each area rated the concepts as "true," "false," "debatable," or "ambiguous." From the jury ratings, approved lists of concepts were established.

Another evaluation by a general education jury established the value of these concepts to general education. A list of comprehensive teachable concepts is available to health educators, science teachers, and others interested in tobacco smoking education.

Introduction

The controversy over the association of and/or the relationship between tobacco smoking, lung cancer, and heart diseases has resulted in confusion, misconceptions, and lack of understanding on the part of health educators, classroom teachers, and others regarding the scientific facts that should be taught youth and adults about this topic. To secure these essential facts is a very difficult task. In searching for the most recent and authoritative facts, a person is likely to become so hopelessly bogged down with conflicting and controversial facts that he may teach the topic superficially or omit it completely.

Most of the states have legislation which requires the teaching of the effects of alcohol, stimulants, and narcotics on the human body. Nicotine, one of the most important ingredients of tobacco, comes under the category of a toxic alkaloid. Yet the material presently being taught in the schools is inadequate and antiquated in terms of currently available information.

School authorities and teachers are constantly confronted with the smoking problem within the school and on the school grounds. Many parents and most of the pupils would like to see smoking permitted within the school building. Some schools have adopted this policy (7).

Information about and authoritative opinions on smoking from medical authorities and health experts are needed (3, 11). The relationship between smoking, especially cigarette smoking, and lung cancer, heart and other cardiovascular diseases needs to be understood. Concepts which should be taught need to be determined and made available especially for teachers of

health, science, and other related fields.

Concept studies have been done in other areas of learning. General science and health (2, 10) are two of the areas. Staton's (10) study in health education concepts was an improvement over some of the earlier studies in that his ideas were rated in terms of their relative importance to education by a jury of experts.

A study by Giovannangeli (4) involved the selection, development, and objective evaluation of concepts in alcohol education. Concepts were secured be reviewing textbooks and professional periodicals. Then they were further separated into three categories of physiological, psychological, and socio-economic. Each concept was rated by a jury of 25 experts as "true," "false," or "debatable." All that were judged to be "true" by 69 percent of the jury members were submitted to a general education jury for further evaluation in terms of their importance to general education.

The term "concepts" used in this study is stated in declarative sentences which represent teachers' goals, recognizable advances in the educative growth to be made by the pupils.

The purposes of this study are (1) to determine the validity and accuracy of tobacco smoking concepts and (2) to determine the tobacco smoking concepts that have value to general education.

Procedure

Selection of the tobacco smoking concepts was done by a thorough examination of secondary school and college health, hygiene, and science textbooks; books about tobacco smoking; numerous pamphlets; medical journals; and other professional and popular sources.

From these many sources an original group of more than 400 concepts was secured. They were critically and thoroughly examined by the authors and translated into more understandable language and terms. Some of them had to be telescoped (1), a technique commonly used in studies of this type in which the concepts are classified and a grouping of ideas which possess similarity is made. This process reduced the number of concepts and was the first attempt to remove ambiguities and duplications to clarify them for better understanding. Obviously, this required some degree of judgment on the part of the authors but extreme care was taken to see that the intended meaning and context were not changed.

For further convenience the concepts were classified into the following areas: (1) physiological effects of tobacco smoking, (2) psychological effects of tobacco smoking, and (3) socio-economic effects of tobacco smoking on the individual.

A preliminary study involved a physician, two health educators, and a physiologist and was undertaken to further clarify and establish the validity of the concepts. Several meetings of this group were held in which ideas were individually examined. Some concepts were reworded to make them

more understandable and valid, others were grouped together, and still others eliminated. General agreement of this subcommittee was secured on the 161 concepts that were retained.

The next step was to select a jury of at least 25 experts in each of the three areas: physiological, psychological, and socio-economic. Care was taken to secure jurors who were truly experts in their representative areas. Twenty-one of the physiological jurors were physicians, most of whom were specialists and many of whom were active in cancer and smoking research. The other three held doctor's degrees in physiology. Psychological jurors were psychologists, in both research and teaching, with doctorates in the field. Sociologists, also with doctorates, made up the socio-economic jury. Jurors for the three areas were selected from all regional areas of the country.

Each juror was asked to rate every concept as "true," "false," "debatable," or "ambiguous." More than 90 percent of those who agreed to be jurors returned the questionnaire. It was decided to make a final listing of rated concepts in order of percent. If all the jurors gave a concept a "true" rating, 100 percent was indicated in the true column. In a like manner, percentages for each of the ratings were determined.

In as controversial an area as cigarette smoking and its association with and/or relationship to health and certain disease conditions it is difficult to secure agreement, even among specialists in the medical profession. In light of this fact and previous studies (4) it was decided to retain any concept that had a 67 percent or more rating of "true" by the jurors. Concepts with this rating were considered to be the valid, approved list of concepts.

Fifty-seven physiological concepts were sent in questionnaire form to 25 physiological jurors. Twenty-four of the 57 physiological concepts were rated "true" by at least 67 percent of the jurors (Table 1). Seventy-five percent of the jurors rated 19 of the concepts "true." Two concepts received a rating of 100 percent. Thirty-seven received at least one "false" rating. Fourteen concepts were rated "debatable" by at least 33 percent of the jury members. Thirty-nine of the concepts received at least one "ambiguous" rating.

Nine of the 24 valid and approved physiological concepts were concerned with the effects of tobacco smoking on the respiratory, digestive, circulatory, and nervous systems. Seven involved the association with and/or relationship to cancer. Individual sensitivity and general bodily reactions to tobacco smoking accounted for another seven of the approved concepts.

Some of the ratings were surprising to the authors. On several occasions concepts taken from the medical journals were given a rating of debatable, ambiguous, or even false by a considerable number of the jurors. A case in point is concept No. 44: *Smoking increases the metabolic rate*. It was rated "true" by 4 percent of the jurors; "false" by 6 percent; "debatable" by 40 percent; and "ambiguous" by 5 percent. One juror did not rate the concept. It was eliminated by these ratings which are surprising in light of the fact

TABLE 1.—EVALUATION OF PHYSIOLOGICAL CONCEPTS FOR TOBACCO SMOKING EDUCATION[a]

	T	F	D	A
1. Smoking is an irritant to the respiratory tract	100	0	0	0
2. Individuals vary in physiological sensitivity to smoking	100	0	0	0
3. Smoking generally reduces the appetite	92	0	8	0
4. Moderate smoking for one person may be excessive for another	92	0	0	8
5. Today many physicians are advising many patients to quit smoking or reduce the number of cigarettes smoked daily	88	0	12	0
6. Only a slightly higher rate of lung cancer is found in pipe and/or cigar smokers than in non-smokers	88	0	4	8
7. Rather violent physiological reactions cause many individuals to become ill after their first experience with smoking	88	0	8	4
8. Body temperatures of the fingers and toes are lowered when one smokes	88	0	8	4
9. Inhaling causes a smoker to absorb more nicotine	83	0	12	4
10. Ulcers of the stomach are more difficult to treat if the patient continues to smoke	83	4	11	0
11. The smoker's cough often masks the first symptoms of lung cancer	83	0	17	0
12. Death rates of heavy cigarette smokers are higher than non-smokers	79	0	8	12
13. The great rise in male lung cancer in the United States is compatible with the increased cigarette consumption	79	0	12	8
14. The physiological effects of smoking on the cardiovascular systems have been demonstrated on the electrocardiogram and the ballistocardiogram	79	0	16	0
15. Smoking affects the circulatory system through the autonomic nervous system	75	8	8	4
16. The incidence of lip and mouth cancer is greater in smokers than in non-smokers	75	0	16	8
17. Most medical authorities believe there is an association between cancer of the lung and cigarette smoking	75	0	24	0
18. Substances taken into the body in the smoke of burning tobacco include nicotine, ammonia gas, pyridine or pyridine derivatives, carbon monoxide, and tobacco tars	75	4	20	0
19. Lung cancer that is caused by tobacco tars is a result of long term exposure to cigarette smoking	75	0	8	16
20. If one must smoke, cigar and pipe smoking have fewer health implications	71	4	20	4
21. The chances of developing lung cancer increases in proportion to the number of cigarettes smoked	71	0	16	8
22. In some individuals smoking causes increased irritability and nervousness	71	4	24	0
23. Heavy smokers past 40 years of age should have an X-ray every six months, preferably every three months	67	4	28	0
24. Heart rate and blood pressure increases due to smoking tend to persist for 10 to 20 minutes after smoking one cigarette	67	4	16	4

[a]Expressed in descending order of percentages, figured to the lowest number. (T—True; F—False; D—Debatable; A—Ambiguous.)

that an increase in metabolism with smoking has been demonstrated by a number of investigations (5, 6, 8).

The rating given concept No. 57 was very interesting: *Recent evidence points to the fact that the increase in the arsenic content of tobacco is related to the increase in lung cancer.* It was rated "true" by 9 percent of the jurors; "false" by 32 percent; and "debatable" by 56 percent,[1] yet this concept was taken from the *New England Medical Journal* (9).

This is an important and frequently ignored problem in determining what concepts should be taught and what materials included in textbooks. No one would question that the rating of expert jurors is preferable to the opinions of one or two individuals, however expert, in determining concepts and subject content, yet here is a weakness of this type of study.

As previously mentioned, the physicians who served as jurors on the physiological jury were specialists. One criterion for selecting these specialists was that they had published an article on smoking or were doing research in this or closely related areas.

It could well be a further challenge to teachers and students using these concepts to obtain local medical opinions on them. In this way, a greater understanding of the reasons for the different opinions could be secured.

Forty-seven psychological concepts were sent to 33 psychological jurors (Table 2). Twenty-four were rated "true" by more than 67 percent of the jurors. Seventeen concepts were rated "true" by more than 75 percent of the jurors; 15, by more than 80 percent; and 6, by more than 90 percent. The concept, *Smoking may become a habit,* was rated "true" by 100 percent of the jurors.

Thirty-seven psychological concepts received at least one "false" rating. Forty-four received at least one "debatable" rating, and 12 were rated "debatable" by more than 33 percent of the jurors. Forty-one received at least one "ambiguous" rating; 13, an "ambiguous" rating by more than five jurors.

The majority of the valid, approved psychological concepts are concerned with the formation of smoking habits, emotional and sensory factors associated with peer acceptance, and feelings of personal and social adequacy.

Thirty-five socio-economic tobacco smoking concepts were submitted for rating to 30 socio-economic jurors (Table 3). Thirteen concepts were rated "true" by more than 67 percent of the jurors; 11, by more than 75 percent; and 5, by more than 90 percent. Two concepts were rated "true" by 100 percent of the socio-economic jurors.

Twenty-four concepts were considered "false" by at least one member of the jury. Twenty of the 35 cencepts received at least one "debatable" rating; ten were rated "debatable" by more than 50 percent of the jury; and 15 by 40 percent or more of the jury members.

---

[1]Some percent ratings do not total 100 percent because they were rounded off to the nearest whole number.

TABLE 2.—EVALUATION OF PSYCHOLOGICAL CONCEPTS FOR TOBACCO SMOKING EDUCATION*

	T	F	D	A
1. Smoking may become a habit	100	0	0	0
2. Adolescents frequently start smoking to indicate adult status	97	3	0	0
3. Sociability, custom, and nervous habits are factors in smoking.	94	3	0	3
4. Some individuals believe it looks sophisticated to smoke	94	0	6	0
5. Some children begin smoking because smoking means breaking away from parental restraint	94	0	3	3
6. A person may smoke because he desires to be on a level with the other fellow	94	0	3	3
7. The desire to smoke is in no way an innate desire	88	3	3	6
8. Religious convictions discourage some people from smoking	88	6	6	0
9. Some people like to smoke for sensory pleasure	88	0	6	6
10. The logical way to decide whether to smoke should involve a consideration of the advantages and disadvantages of smoking	85	6	6	3
11. In a time of crisis, the smoker may find a quieting power in a cigarette	85	6	3	6
12. Many individuals smoke for relaxation	85	3	12	0
13. The fragrance of tobacco is appealing to some smokers	85	3	12	0
14. Many individuals smoke because their associates smoke	82	3	12	3
15. A strong motivation must be present to break the smoking habit	82	0	6	12
16. People smoke to relieve tension	79	0	12	9
17. Many individuals feel a need to do something with themselves and their hands so they smoke	76	0	18	6
18. Smoking habits may be changed by psychologically oriented advertising	73	6	18	3
19. Some people smoke to relieve anticipated stress	73	0	24	3
20. Individuals smoke to express sociability and an aid to poise	73	3	15	9
21. Smokers are stimulated by watching others smoke	73	3	18	6
22. Many individuals do not smoke because of participation in athletics	70	9	15	6
23. Fear of certain disease implications associated with smoking is a deterrent to some potential smokers	67	9	18	3
24. The feel of the cigarette in their lips is pleasant to many	67	6	24	0

*Expressed in descending order of percentages.
(T—True; F—False; D—Debatable; A—Ambiguous.)

The authors considered the list of socio-economic concepts submitted to the jurors to be equally as good a list as the physiological and psychological concepts. The high percentage of "debatable" and "ambiguous" ratings was unexpected and difficult to understand. Since all the concepts were taken from reliable sources, the authors were forced to conclude that many of the socio-economic jurors were not familiar with some of the concepts presented. It was not thought necessary to include a "don't know" rating for the jurors, but undoubtedly the "debatable" and "ambiguous" ratings frequently were given in those circumstances. In relation to health education teaching, the socio-

TABLE 3.—RATING OF SOCIO-ECONOMIC CONCEPTS FOR TOBACCO SMOKING EDUCATION[a]

	T	F	D	A
1. Smoke from tobacco may be annoying and unpleasant to other persons	100	0	0	0
2. Some employers do not like employees to smoke on the job	100	0	0	0
3. Some individuals feel that smoking is a social requirement	97	0	3	0
4. Advertising and merchandising techniques helped break down the prejudice of cigarette smoking by women	90	0	10	0
5. Tobacco manufacture is among the top 20 industries in national advertising	90	0	10	0
6. In the last few years the sale of king-size cigarettes has increased rapidly	83	0	17	0
7. Tobacco is an important cash crop in American farming	83	3	6	6
8. The great increase in smoking by women in recent years is due, in part, to the breakdown in prejudice against cigarette smoking	77	3	6	13
9. The sale of filter tip cigarettes is constantly increasing	77	0	23	0
10. The great majority of tobacco tax receipts comes from the sale of cigarettes	77	0	22	0
11. Most male smokers are cigarette smokers	77	3	20	0
12. The truthfulness of cigarette advertising has frequently been questioned by the Pure Food and Drug Administration	70	3	20	3
13. Approximately 80 percent of all tobacco produced in the United States goes into cigarettes	67	0	33	0

[a]Based on percent of rating by 30 jury members.
(T—True; F—False; D—Debatable; A—Ambiguous.)

economic concepts as determined by the general education jury are of less importance than the physiological and psychological ones.

The juries' determination of 24 physiological, 24 psychological, and 13 socio-economic concepts was considered to be the most valid and approved list of tobacco smoking concepts, 61 in all. They were considered "true" by 67 percent or more of the experts.

A second purpose of this study was to have the physiological, psychological, and socio-economic concepts as determined by the subject matter expert jurors submitted for further evaluation to a general education jury. It is quite possible that certain concepts would be considered important by the subject matter experts and not feasible or practical to be included in the general education program for all students.

A jury of 25 general educators, college professors in the areas of general education throughout the country, was selected to make the evaluation. Each of the members was asked to evaluate the concepts as (1) very important to general education, (2) moderately important to general education, (3) slightly important to general education, and (4) of no value to general education. In order to equalize the evaluation the following numerical values were given to the ratings: 4—very important; 3—moderately important; 2—slightly important; 1—not important.

The rank order of each concept was determined by adding up the total of these responses in terms of their assigned numerical values. The highest possible score with 25 jury members was 100 points and the lowest 25 points.

All concepts which had a numerical value of 76 or above (76 percent) were considered very important to general education; those from 50-75, moderately important; those from 25-49, slightly important; and below 25, of no importance to general education.

The general education jurors rated 14 of the 24 physiological concepts (Table 4) as very important to general education. The remaining ten physiological concepts were considered moderately important. All the physiological concepts were rated as very important or moderately important to general education.

Three of the 24 psychological concepts were considered very important to general education; 18, moderately important; and three slightly important. None of the psychological concepts was considered of no importance.

The general education jury considered two of the 13 socio-economic concepts very important to general education. Eight concepts were of moderate importance and three of slight importance to general education.

Considering the 61 concepts as a group (physiological, psychological, socio-economic), the general education jury rated 19 very important to general education, 36 moderately important, and six slightly important, but none of the concepts of no importance.

According to the general education jurors, the main emphasis in tobacco smoking education should be on the physiological concepts (Table 4), including the medical and health implications associated with tobacco smoking which everyone should understand. They considered the psychological concepts of secondary importance and the socio-economic concepts of minor importance to general education.

Summary and Conclusions

Physiological, psychological, and socio-economic tobacco smoking concepts to be used in health education teaching were secured from many sources. They were translated, telescoped, and refined to make them more valid and understandable.

These concepts (161 in all areas) were submitted to expert jurors (25 or more) in each area who rated the concepts as "true," "false," "debatable," or "ambiguous." The rank for each was obtained by listing them in order of percent for each rating. Those which had a "true" rating higher than 67 percent were considered the most valid and approved list of concepts for tobacco smoking education.

A second rating to determine the importance of the concepts to general education was undertaken. An expert jury of general educators was selected to rate the concepts: 24 physiological, 24 psychological, and 13 socio-economic. They rated the concepts as very important to general education,

TABLE 4.—GENERAL EDUCATION JURY RATINGS OF THE IMPORTANCE OF PHYSIOLOGICAL, PSYCHOLOGICAL, AND SOCIO-ECONOMIC TOBACCO SMOKING CONCEPTS[a]

Very Important to General Education

Physiological

1. Most medical authorities believe there is an association between cancer of the lung and cigarette smoking.
2. Smoking is an irritant to the respiratory tract.
3. Death rates of heavy cigarette smokers are higher than non-smokers.
4. The chances of developing lung cancer increase in proportion to the number of cigarettes smoked.
5. Inhaling causes a smoker to absorb more nicotine.
6. The great rise in male lung cancer in the U. S. is compatible with the increased cigarettes smoked daily.
7. Today many physicians are advising many patients to quit smoking or reduce the number of cigarettes smoked daily.
8. The physiological effects of smoking on the cardiovascular system have been demonstrated on the electrocardiogram and ballistocardiogram.
9. Lung cancer that is caused by tobacco tars is a result of long-term exposure to cigarette smoking.
10. Individuals vary in physiological sensitivity to smoking.
11. Moderate smoking for one person may be excessive for another.
12. The incidence of lip and mouth cancer is greater in smokers than in non-smokers.
13. Substances taken into the body in the smoke of burning tobacco include nicotine, ammonia gas, pyridine or pyridine derivatives, carbon monoxide, and tobacco tars.
14. Heavy smokers past 40 years of age should have an X-ray every six months.

Psychological

1. Smoking may become a habit.
2. A strong motivation must be present to break the smoking habit.
3. Adolescents frequently start smoking to indicate adult status.

Socio-Economic

1. The truthfulness of cigarette advertising has frequently been questioned by the Pure Food and Drug Administration.
2. Smoke from tobacco may be unpleasant and annoying to other persons.

Moderately Important to General Education

Physiological

1. In some individuals, smoking causes increased irritability and nervousness.
2. Smoking affects the circulatory system through the autonomic nervous system.
3. Ulcers of the stomach are more difficult to treat if the patient continues to smoke.
4. The smoker's cough often masks the first symptoms of lung cancer.
5. Only a slightly higher rate of lung cancer is found in pipe and/or cigar smokers than in non-smokers.
6. Heart rate and blood pressure increases due to smoking tend to persist for 10 to 20 minutes after smoking one cigarette.
8. Smoking generally reduces the appetite.
9. Rather violent physiological reactions cause many individuals to become ill after their first experience with smoking.
10. Body temperature of fingers and toes are lowered when one smokes.

[a] Ratings of 25 experts in general education.

Psychological

1. The logical way to decide whether to smoke should involve a consideration of the advantages and disadvantages of smoking.
2. Sociability, custom, and nervous habit are factors in smoking.
3. Smoking habits may be changed by psychologically oriented advertising.
4. Fear of certain disease implications associated with smoking is a deterrent to some potential smokers.
5. A person may smoke because he desires to be on a level with the other fellow.
6. Some children begin smoking because smoking means breaking away from parental restraint.
7. Many individuals smoke because their associates smoke.
8. Some individuals believe it looks sophisticated to smoke.
9. Some people smoke to relieve anticipated stress.
10. Many individuals do not smoke because of participation in athletics.
11. People smoke to relieve tension.
12. Individuals smoke to express sociability and as an aid to poise.
13. Many individuals smoke for relaxation.
14. The desire to smoke is in no way an innate desire.
15. In a time of crisis, a smoker may find a quieting power in a cigarette.
16. Many individuals feel a need to do something with themselves and with their hands so they smoke.
17. Some people like to smoke for sensory pleasure.
18. Smokers are stimulated by watching others smoke.

Socio-Economic

1. Some employers do not like employees to smoke on the job.
2. Advertising and merchandising techniques helped break down the prejudice of cigarette smoking by women.
3. Tobacco manufacture is among the top 20 industries in national advertising.
4. Approximately 80 percent of all tobacco produced in the United States goes into cigarettes.
5. Some individuals feel smoking is a social requirement.
6. The great increase in smoking by women in recent years is due, in part, to the breakdown in prejudice against cigarette smoking.
7. Tobacco is an important cash crop in American farming.
8. The great majority of tobacco tax receipts comes from the sale of cigarettes.

Slightly Important to General Education

Physiological

None.

Psychological

1. Religious convictions discourage some people from smoking.
2. The fragrance of tobacco is appealing to some smokers.
3. The feel of the cigarette in their lips is pleasant to many.

Socio-Economic

1. The sale of filter tip cigarettes is constantly increasing.
2. Most male smokers are cigarette smokers.
3. In the last few years the sale of king-size cigarettes has increased rapidly.

Of No Importance to General Education

None

moderately important to general education, slightly important, or of no importance to general education.

In the final rating, 19 concepts, 14 of them physiological, were considered very important to general education; 36, moderately important; and 6, slightly important.

A list of teachable concepts for tobacco smoking education was established by the rating of subject matter specialists in the physiological, psychological, and socio-economic areas. A further rating of their value to general education was made by general educators and education jurors. To the authors' knowledge, this is the only comprehensive list of concepts available for health educators, science teachers, and others interested in tobacco smoking education.

References

1. Charte, W. W., and Waples, D. *The Commonwealth Teacher Training Study.* Chicago: University of Chicago Press, 1929.
2. Craig, Gerald S. "Certain Techniques Used in Developing a Course of Study in Science for the Horace Mann Elementary School." New York: Bureau of Publications, Teachers College, Columbia Univ., 1927.
3. Galarneaux, Alexander F. *The Selection, Development and Objective Evaluation of Tobacco Smoking Concepts as an Aid to Curriculum Construction.* Unpublished Doctoral dissertation. Boston: Boston Univ., 1957.
4. Giovannangeli, Arthur J. *The Selection, Development, and Objective Evaluation of Concepts in Alcohol Education as an Aid to Curriculum Construction.* Unpublished Doctoral dissertation. Boston: Boston Univ., 1954.
5. Goddard, V. R., and Voss, J. C. "The Immediate Effects of Cigarette Smoking upon Basal Metabolism of University Men and Women." *Journal of Laboratory and Clinical Medicine* **27**: 787-91; Mar. 1942.
6. Hiestand, W. A.; Ramsey, H. J.; and Hale, D. M. "The Effect of Cigarette Smoking on Metabolic Rate, Heart Rate, Oxygen, Pulse, and Breathing Rate.' *Journal of Laboratory and Clinical Medicine* **25**: 1013-17; Mar. 1940.
7. Patterson, Walter G. *Smoking in High; Report of Ten Years Working with the Smoking Problem.* Bulletin No. 40. Washington, D. C.: National Association of Secondary-School Principals, a department of the National Education Association, 1956.
8. Roth, G. M.; McDonald, J.; and Sheard, B. "The Effects of Cigarette Smoking." *Journal of American Medical Association* **11**: 761-67; July 15, 1944.
9. Satterlee, H. S. "The Problem of Arsenic in American Cigarette Tobacco." *New England Journal of Medicine* **25**: 1149-55; June 21, 1956.
10. Staton, Wesley M. *A Determination of Fundamental Concepts of Healthful Living and Their Relative Importance for General Education at the Secondary Level.* Unpublished Doctoral dissertation. Boston: Boston Univ., 1948.
11. Thompson, Clem W., and Flynn, W. "New Health Education Implications in the Association of Smoking, Lung Cancer and Heart Disease." *Research Quarterly* **27**: 431-45; Dec. 1956.

(Submitted 8/20/58)

Influence of Measurement Error and Intra-Individual Variation on the Reliability of Muscle Strength and Vertical Jump Tests[1]

FRANKLIN M. HENRY
University of California
Berkeley, California

Abstract

Independent estimates of the errors of measurement and the intra-individual variations have been made for arm strength (Smedley dynamometer) and vertical jumping (Henry apparatus). The intra-individual variations are much larger than the measurement errors; thus they constitute the chief factor that determines test-retest reliability in these two physical performances. The tolerable error for measuring strength is ± 0.82 kg.; for jumping it is ± 0.5 in. In order to avoid loss of accuracy, it is necessary to read the dial or scale with approximately twice this precision. The method of computing test-retest reliability as the ratio of "true score" variance to total variance is found to underestimate the coefficient when the variability of test and retest scores differs by more than 15 percent. A formula for correcting this attenuation is presented.

THE TEST-RETEST reliability coefficient can be visualized and calculated as the ratio of "true" individual difference variance to total variance, using the same data that are required for the conventional correlation method (2). If an independent estimate of the "true" error of measurement is available, this ratio calculation permits evaluation of the separate roles of measurement error and intra-individual variation, as determiners of the size of the reliability coefficient. The method has been used in another study to interpret the factors determining reliability in typical reaction and movement time data (2). The present investigation is concerned with analyzing the reliability of a strength test and a vertical jump test.

Dynamometer Error

Ordinary calibration methods, while useful for estimating the *constant* error of an instrument such as the Smedley hand dynamometer, do not furnish information on the *variable* error. However, it is the latter, rather than the former, that is important in test reliability.

We have available in this laboratory a mechanical recording device for the Smedley instrument which has been described in another report (1). It

[1]From the Research Laboratory of the Department of Physical Education. The writer is indebted to Dollie Carston for technical assistance in obtaining the dynamometer scores.

has been subsequently modified by attaching the dynamometer directly to the paper drive unit (which is now mounted on gimbals), so that uncertainty in the flexible coupling is considerably reduced. Using this recording instrument, one can read in the usual manner the maximum deflection of the gear-driven pointer attached to the Smedley dial, and also measure the deflection of the recording pen which is directly connected to the movable handle of the dynamometer by a flexible cable. The squared standard deviation of the discrepancies between the two readings in a series of tests is the measurement error variance (σ_e^2) of the *sum* of the uncertainties of the two indicating systems including the errors of reading. The observed total error must therefore be larger than that for each system considered separately. Nevertheless, its total amount is relatively small, as will be shown.

Role of Measurement Error in Strength Reliability

Table 1 summarizes the results of 41 test-retest strength determinations on college men and 33 test-retest determinations on college women. The tests and retests were separated by approximately one week. In each case, the maximum pen deflection in a series of contractions has been used—usually this was the first squeeze, although it was the second in 6.1 percent of the tests, the third in 4.1 percent, and the fourth or later in 3.4 percent. The dynamometer dial was read to the nearest 0.5 kg.; the pen recording to the nearest 0.2 kg.

In Table 1, the total variance σ_x^2 is the average of the test and retest variances σ_a^2 and σ_b^2, the intra-individual variance σ_i^2 is $\frac{1}{2}$ (σ^2_{a-b})—σ^2_e; the

TABLE 1.—INFLUENCE OF MEASUREMENT ERROR AND INTRA-INDIVIDUAL VARIATION ON RELIABILITY

Statistic		Dynamometer Strength (kg.)		Vertical Jumping (cm.)	
		Men	Women	Single Trial	Five Trial
Test A	M_a	60.850	36.882	48.506	48.695
	σ_a	5.358	6.423	8.200	7.651
Test B	M_b	60.231	37.000	49.862	49.128
	σ_b	4.930	6.714	8.505	7.982
Total variance	σ_x^2	26.507	43.166	69.788	61.125
Between day variance	$\sigma^2_{a-b}/2$	6.553	5.333	10.158	1.582
Measurement error	σ_e^2	0.984	0.984	0.084	0.017
Intra-individual	σ_i^2	5.569	4.349	10.074	7.565
Inter-individual	σ_t^2	19.954	37.833	59.630	59.543
Test-retest reliability	r_x	.753	.876	.854	.974
Corrected reliability	r_i	.782	.897	.855	.974

"true score" or inter-individual difference variance σ_t^2 is $\sigma_x^2 - \sigma_i^2 - \sigma_e^2$. The reliability coefficient is the ratio σ_t^2/σ_x^2, as explained in the earlier article (2). The influence of the measurement error on the reliability coefficient can be removed by subtracting σ_e^2 from σ_x^2 in the above ratio. (Note that it has already been removed from σ_t^2). This corrected coefficient has been designated r_i to emphasize that the only reason it is less than 1.00 is the test-retest variability caused by the intra-individual variance σ_i^2.

It may be calculated from the tabled data that the measurement error accounts for 15.0 percent of the observed between-day variance in the mean; its removal increases the reliability from .753 to .782, a gain of .029. For the women, the measurement error is 18.4 percent; removing it increases the reliability from .876 to .897, a gain of .021. While the effect is noticeable, it is certainly not large. As mentioned earlier, the true error variance was probably about half as large as the tabled value; if this estimate is accepted, the gains are only .014 and .010.

Jumping Apparatus Error

The vertical jumping recorder consisted of a cord attached to the subject's head, passing upward to a ceiling pulley and then downward to a sliding indicator and a tension reel. It was described in a recent article by Pacheco (4). In order to create a reproduceable controlled movement equivalent to a jump, the subject's cord was brought down vertically from the ceiling pulley to form a loop near the floor, and then upward a few feet to a tie point on the wall. The bottom of the loop passed over a 6 in. grooved pulley mounted on a 15 in. lever that was firmly pivoted on the wall. A handle was attached to the lever at the pulley bearing. With normal tension on the cord, swinging the lever from its highest to its lowest position about its own radius produced a vertical cord movement that was first positively and then negatively accelerated. This closely approximated an actual jump. The standard deviation of 100 trials has been used to estimate the measurement error variance σ_e^2 that is given in Table 1.

Role of Measurement Error in Jump Test Reliability

The data of the Pacheco Experiment 2, consisting of five trials for each of 50 college men on two days, separated by a week, are available (4). From the raw data,[2] trial #5 on each of the two days has been used to compute the single trial reliability. The results are given in Table 1. The error vari-

[2]Betty A. Pacheco, personal communication. Attention is called to a small but important omission in the description of the jumping apparatus (4). As the subject goes down into his crouch before he jumps, the sliding indicator must be quickly pushed to the right so that it goes close to the minimum position of the driving slider. Otherwise, the indicator may bounce on the upward part of the jump and read too high. It should also be mentioned that in using this apparatus, we have consistently employed a *modified* Sargent jump, with the downward arm snap eliminated under the assumption that it might introduce a learning factor.

ance has of course been divided by five to compute σ_e^2 for the five-trial data.

In this experiment, the measurement error variance constitutes only 0.83 percent of the single trial variance between days, and only 1.07 percent in the case of the five-trial averages. In consequence, there is no appreciable change in the reliability coefficient when the influence of errors of measurement is removed.

As a matter of interest, a calculation has been made of the amount of measurement error that would be appreciable although still tolerable, arbitrarily taken as a decrease from .855 (corrected) to .835. Since σ_t^2 would remain unchanged, σ_e^2 would be 1.67 cm.; in other words, the tolerable error (σ_e) would be $\pm$ 1.29 cm. or approximately a half-inch. Error of this magnitude would only reduce the five-trial reliability by .005, which is a trivial amount. It seems, therefore, that the measurement error in jumping performance, regardless of the instrument used, is tolerable if kept within a half-inch.

Agreement Between Reliability Methods

It has been observed in these studies that the size of the reliability coefficient is the same, whether computed as σ_t^2/σ_x^2 or by the conventional method, provided that σ_a^2 and σ_b^2 are approximately equal. However, when the difference between the test and retest variances is as large as 15 to 20 percent, the present method gives a value that is about .003 correlation units too low. The agreement can be made almost exact if σ_x^2 is empirically corrected by adding to it the factor 0.0643 $(\sigma_a^2 - \sigma_b^2)$, using the corrected value of σ_x^2 and the *uncorrected* between-day variance for computing σ_t^2 in the ratio σ_t^2/σ_x^2.

Discussion and Conclusions

It is clear that if one is willing to accept a decrement of .02 correlation units in the reliability coefficient as reflecting a reasonable amount of measurement error, the tolerable amounts of variable error (σ_e) are $\pm$ 0.82 kg. for the hand dynamometer, $\pm$ 0.5 in. for the vertical jump, and (as reported in reference 2) $\pm$ 0.006 sec. for reaction or movement times. These generalizations of course apply only to groups of subjects whose variability is comparable to the samples investigated.

It is also clear that in these three tests at least, the important source of unreliability is *not* the measurements in the testing. It is the *intra-individual* variation of the subjects who are tested. Nevertheless, it *is* necessary to know that a particular hand dynamometer is consistent to the required extent, and, furthermore, it must be read to within the nearest half kg. or errors in reading will have an appreciable influence. This is shown by the fact that the between-day variance increases by 9.2 percent when the raw dynamometer scores are rounded off to the nearest whole kg. unit. This is equivalent to introducing an additional error variance of 0.5 kg. Similarly, hundredth second times

must be read to within a half-hundredth in reaction time measurements (3). Using the same standard (7.5% of the between-day variance), the height jumped should be measured to the nearest 5/16 inch (0.8 cm.).

The writer hopes that others will become interested in applying this type of analysis to different tests commonly used in our field. It is quite possible that we insist on unduly expensive measurement accuracy in some situations where we do not need it (because of limitations imposed by the intra-individual variation). At the same time, we may be blissfully unaware of the need for more refined measurement in certain other situations.

References

1. Grose, J. E. "Depression of Muscle Fatigue Curves by Heat and Cold." *Research Quarterly* **29**: 19-31; Mar. 1958.
2. Henry, F. M. "Reliability, Measurement Error, and Intra-Individual Differences." *Research Quarterly* **30**: 21; Mar. 1959.
3. ———. "Precision and Operation of the Hundredth-Second Electric Timer." *Research Quarterly* **30**: 117; Mar. 1959.
4. Pacheco, Betty A. "Improvement in Jumping Performance due to Preliminary Exercise." *Research Quarterly* **28**: 55-63; Mar. 1957.

(Submitted 8/28/58)

Contributions of R. Tait McKenzie to Modern Concepts of Physical Education

ADELAIDE HUNTER
Pan American College
Edinburg, Texas

Abstract

This historical study reveals the educational leadership of R. Tait McKenzie in the development of some modern concepts of physical education. A summary of the major theories of Dr. McKenzie concerning the role of physical education in daily living, the college curriculum, professional standards, teacher education, and the place and organization of athletics in the educational program is presented.

Introduction

The purposes of this study are to reveal how one individual contributed to the enrichment of the profession and to indicate some of the important concepts which influenced the development of physical education. Although the main outlines of the history of physical education are generally known, many details of individual accomplishment are needed to complete the account. Studies of this type may serve as a source of inspiration for present and future members of the profession as well as a basis for better understanding of physical education by those outside the profession.

This study developed as a result of a more extensive research problem concerning the total contributions of R. Tait McKenzie to the field of physical education (3). The analysis of information obtained from pertinent literature, letters, and personal interviews demonstrated that Dr. McKenzie had a profound influence upon the development of modern concepts in a rapidly growing profession.

No attempt is made in this study to discuss Dr. McKenzie's educational philosophy in detail or to state his influence upon all aspects of modern physical education. The following account is a summary of the contributions of R. Tait McKenzie to the major concepts of present-day physical education.

AT THE CLOSE of the first two decades of the twentieth century, the traditional aims of education in the United States were gradually replaced by aims which were based upon scientific psychology and social philosophy (1, p. 643). These general educational changes were reflected in the field of physical education. The beginning of the 1920's saw the widening of professional horizons to include social and psychological objectives at the expense of those that had been mainly corrective and disciplinary in nature.

One of the forces that aided in this enrichment of the objectives of physical education was the work of a committee of the Society of Directors of Physical Education in Colleges. This committee was appointed to formulate the aims

and scope of physical education, and the persons charged with this responsibility were Fred E. Leonard, Joseph E. Raycroft, and R. Tait McKenzie.

The report of this committee (4) set forth a definition of physical education which called for an evaluation of the current aims, methods, and practices in the field. This definition stated that:

> The term "physical education" should be conceived as being in conformity with the present conception of man's nature as a unit, and which sees in measures insuring bodily health and the right kind and amount of motor activity an avenue of approach through which the whole individual may be influenced for good, in mind and in character as well as in body; it employs the word physical to denote the means, not the end.

It is interesting to note that, in defining the aims of physical education, the committee formulated fundamental concepts that differ very little from those advocated by physical educators today. The first of these aims was concerned with social significance. The committee believed that:

> If we conceive the perfecting of the individual in his social relations to be of greater importance than more purely personal values, we may begin our list of aims with certain qualities developed by appropriate group activities, particularly athletic games and sports, practiced under favorable conditions. It is through these agencies that the child and youth most readily and naturally acquire habits and attitudes of cooperation and friendliness, loyalty, capacity for leadership, a spirit of fair play, and all that is implied in the word sportsmanship.

The second aim of the committee stated that physical educators should strive to realize that the developing of worthwhile traits was of both community and individual significance. The committee further believed that such traits of self-confidence, alertness, resourcefulness, decision, perseverance, courage, aggressiveness, and initiative must be ensured through other means than those which the average family life could provide.

The committee contended that the underlying purposes of these aims were the following: to promote the normal growth and organic development of the individual, to conserve his health, to provide a maximum degree of efficiency and endurance, and to secure the neuro-muscular control required for graceful and effective movements. The final aim for physical education, as established by this committee, was: "To engender in youth an intelligent and healthful interest that shall lead to lifelong practices of forms of active exercise."

Dr. McKenzie believed that the object of education is to develop and direct the potentialities of an individual in such a way as to make him most useful as a citizen, to himself and to society. He was of the opinion that each person should be directed into his avenue of optimum development and at the same time should fulfill the requirements of responsible citizenship.

The translation of these aims into the actuality of a college program is suggested by Dr. McKenzie as follows:

> The rounded college course in physical education should then include those formal gymnastics and the gymnastic games that train the body to know its possibilities and

limitations[1] . . . . It should include those athletic sports for all but the physically unfit that cultivate individual daring, courage, and pluck; and it should not overlook those games which are the epitome of life, where inherent manliness is put to the proof, and where, as representatives of their club or college, they have to uphold her good name on track and field, as in future life they will be called upon to do by their town, party, or country (7, p. 779).

The interpretation of these beliefs was carried out through the physical education program which McKenzie directed at the University of Pennsylvania. A varied program of designed and graded activities was offered for both men and women in addition to a comprehensive selection of intramural and intercollegiate athletics. The basis of this program was the physical examination, a novel idea which was to be accepted later as a fundamental practice in physical education. Those students who were found to be physically unable to participate in the regular program received instruction in a special course which placed emphasis upon corrective and adapted activities. The students who were physically fit, and who possessed an average degree of skill, were placed in the regular program; those who were highly skilled participated in the intramural and intercollegiate athletic programs in addition to regular instruction (8, p. 21).

All students who were physically able were required to learn to swim. Once this requirement had been met and competency demonstrated in the fundamental physical skills, the students were allowed a wide latitude in the selection of courses. This means of recognizing individual needs, differences, and interests approximated an ideal program of physical education. Directors of physical education in colleges came from all sections of the United States to see this pioneer program in operation (2).

Dr. McKenzie's belief in physical competency for all students was exceedingly strong. He saw this quality as a basic requirement for the maximum development of the individual in his life activities. He emphasized this belief in the following manner:

The cultivation of physical intelligence can never lose its value, no matter how artificial may be the conditions of the community in which one lives. It is what teaches a man to meet the many emergencies of daily life (6, p. 252).

Dr. McKenzie thought that since physical competency was of such significance in everyday living, the standards of skill in performance should be superior. He further emphasized this view when he said:

---

[1]Dr. McKenzie explained his definition of gymnastics when he said: "All teaching, from the simplest exercise to the most complicated coaching of a team is 'gymnastic' in character according to the correct definition of the term. . . . The boxer in preparing for a fight punches the bag, he shadow-boxes, he practices certain leads and counters again and again in order to perfect his speed and coordination. These are gymnastic exercises. . . . A good coach of basketball puts his squad through graded progressive instruction in dribbling, pivoting, goal shooting, and passing. These are gymnastic drills. . . ."

Standards of performance . . . should be set so that about 50% of the students examined will pass and the other 50% will fail. Those who pass may go on to higher things and will be given a wider selection of what they may do, but the "dubs" should be taught until they can pass them; and this will always be the hardest and least spectacular work of the department. . . . The teacher's success, however, will be in proportion to his ability to analyze games, sports, gymnastics, and acrobatic exercises in such a way to awaken the student's interest in them; but whether the student is enthusiastic or not, he learns to do certain things that he ought to know. A man may try to get out of learning to swim, and many do, but when he has learned to swim, at least he can swim (11, p. 88).

In spite of this seemingly rigid view in regard to the development of physical competency, Dr. McKenzie recognized that any attempt to make an individual conform to an average by enforcing restrictions or by changing his environment was doomed to failure (12, p. 342). He further recognized that any course in physical education must always take into account individual needs and interests and the characteristics of physical growth (6, p. 76). He considered the total development of the individual, however, and believed that it could not be complete without maximum realization of physical potentialities.

During his career as a physical educator, R. Tait McKenzie was constantly endeavoring to elevate physical education to full educational status. He attempted to raise the standards of professional performance and to erase the stigma that had been placed upon physical education by those without regard for scientific principles. The convictions that guided Dr. McKenzie's actions throughout the struggle for academic recognition of physical education not only serve as a source of inspiration for the physical educator but also tend to summarize the basic aspects of his role in the educational scheme. These views are presented in the following words of McKenzie:

If there is one thing that we need more than another it is the ability to drive ourselves to do the hard or disagreeable thing because it is the best and right thing to do—to see the ultimate gain rather than the immediate loss of comfort or ease; and as educators we are too often apt to take the easy way; to accept the plausible excuse; to condone the sloppy performance; to evade the irksome insistence on learning the difficult feat; or to shirk the teaching of the student to overcome his own inertia or fear. . . . Docility and regularity of attendance are not enough to qualify a student for credit in physical education, nor is it enough for a coach to spend his time and energy in securing ready-made material for his teams instead of developing skill in students that are already under his care. If physical education degenerates to the point where the teacher only takes the roll and keeps order, he descends to the level of a policeman and not a teacher. . . . We must show progress in actual teaching all the way from the supervised work of the defective through the mass of the indifferent and non-athletic up to the skilled coaching of men who represent their college in intercollegiate competition. Unless we can do all this, we have no right to a place at the academic fireside (11, p. 89).

In keeping with these convictions, Dr. McKenzie worked energetically throughout his career for the improvement of the status of teacher education. It was his belief that every physical educator should have a broad and catholic background, that he should be required to meet standards of professional competency, and that he should be qualified for academic rank on a basis of

equality with faculty members of other subjects. Concerning this aspect Dr. McKenzie wrote:

> The education of teachers is a function of the program of physical education, and here I believe that a great deal of responsibility exists. If physical education is to maintain its position as a part of the regular college work, it will depend on the education of those who are in charge of it, so every university ought to be prepared to give courses to such students as may elect them (5, p. 36).

Dr. McKenzie endeavored to inspire members of the profession to maintain ideals in teacher education which were consistent with the total educational program. He also worked for the advancement of physical education as an indispensable part of modern living. During his period of professional service he helped in the formulation of principles, aims, and procedures for the enrichment of physical education as a profession and for its expansion as a vital part of the school and college curriculum.

From the very beginning of his career in physical education, R. Tait McKenzie had the vision to see the educational implications of intercollegiate athletics and their powerful influence on the lives of youth. He affirmed our belief in the potential value of athletics when he said:

> Certainly athletics in some form will continue as a college institution so long as a young man's glory is in his strength. He will devise some method of measuring it with his fellows, and of displaying his prowess, even if it be at the expense of the unfortunate policeman or the innocent street lamp. He must have some safety-valve to let off his surplus vital force.
>
> Haphazard as their regulation is, athletic sports have had a powerful influence in molding the lives of men . . . and if the authorized physical education department does not recognize the "play instinct," it will manifest itself as a discordant factor, opposing where it should cooperate, interfering with and detracting from the popularity of the official work (10, p. 24, 26).

McKenzie did not believe, however, in the promotion of intensive intercollegiate athletics at the expense of the average college student who would probably never become a great athlete. He also took a firm stand against the extension of special privileges to athletes. At the same time, he asserted that the coach of competitive sports should be a qualified member of the institutional faculty and primarily a full-fledged educator. In relation to the position which the coach should hold in the educational curriculum, Dr. McKenzie stated:

> The moral influence of the coach for good or ill on the students under his direction must always be overwhelming. He is the center of hero worship, and if he is a "rough-neck," this is speedily known in other colleges and the institution he represents is rated accordingly, and to a large extent justly; and its intercollegiate relationships suffer in consequence (9, p. 12).

Dr. McKenzie believed that the primary aim of intercollegiate athletics should be educational and that this division of physical education should be under the full control of the educational institution concerned. His struggle

against the evils that overshadowed the educational values was a continuous one manifesting itself in many forms. One of the major objectives of the physical education program which McKenzie administered was to instill the ideals of sportsmanship and fair play in the mind and heart of every boy who represented his club or college upon the playing field. His consideration for rival institutions set an example for all to follow in the extension of courtesies to the representative teams. He firmly insisted that athletes be the personification of the amateur spirit. He set up an integrated program of intramural athletics at the University of Pennsylvania which served as a model for many educational institutions, and in addition devoted a great part of his time and effort to the formulation of honor codes, rules, and general principles for the conduct of intercollegiate athletics.

Some of the most important contributions that Dr. McKenzie made to intercollegiate athletics were his efforts to change the influence of the alumni on athletics at the University of Pennsylvania, to reduce the evils of over-organization, and to drive out professionalism. These problems which McKenzie faced were local ones in nature with national significance and implication.

For a period of almost 30 years, Dr. McKenzie waged a one-man struggle against the Council of Athletics which was independent of the University of Pennsylvania, but in full control of the athletic program. The final outcome of this controversy was the adoption of the Gates Plan in 1931. This plan established a Department of Physical Education in the University of Pennsylvania, headed by a dean who was responsible only to the president. The department was set up to constitute three divisions, each headed by a director who was responsible to the dean. The three divisions were the Division of Student Health, the Division of Physical Instruction, and the Division of Intercollegiate Athletics. Other changes were also made in the revision of intercollegiate relations through the Gates Plan. The distribution of financial aid to students was the work of the University welfare committee; eligibility was determined by a faculty committee; the budget was prepared by the dean; and the treasurer of the University supervised financial expenditures. Thus, in the words of Dr. McKenzie, "Sports were given back to the students, teaching to the Faculty, and the deficit to the Treasurer" (8, p. 25). This pioneer development at a leading university was an important advance toward the achievement of the avowed objectives of athletics.

When Dr. McKenzie retired as director of the Physical Education Department at the University of Pennsylvania in 1931 to accept the chair of research professor of physical education, he could look back upon the results of his labors with a great deal of satisfaction. The first class of the undergraduate school for physical education majors graduated in 1928, and the school gave evidence of continued growth and development; there had been a complete revision of intercollegiate athletic relationships with the abolishment of the training house, the curtailment of football training, and the selection of coaches of faculty caliber; there had been a great expansion in facilities and

equipment; and the regular program of physical education reflected the wholesome nourishment and the wise guidance that it received from R. Tait McKenzie (13, p. 132).

The contributions of Dr. McKenzie as a physical educator were numerous and varied. His professional life presents a record of long and active service and upon this record are written ideals and ideas which are as basically sound today as they were revolutionary in his early career. In the final analysis, the nature of his educational leadership is most effectively expressed by the members of the American Association for Health, Physical Education, and Recreation in their presentation of the Fellowship Award to Dr. McKenzie for distinguished service to the profession:

> Ever untiring, ever standing for the highest ideals of life, Dr. McKenzie has moulded in the hearts and souls of an army of college men, the spirit of joy and beauty he so effectively portrayed in sculptors' clay (14, p. 17).

References

1. Butts, Freeman R. *A Cultural History of Education.* New York: McGraw-Hill Book Co., 1947.
2. Hollenback, William. Interview with author.
3. Hunter, Adelaide M. *R. Tait McKenzie: Pioneer in Physical Education.* Unpublished Doctoral dissertation. New York: Teachers College, Columbia University, 1950.
4. Leonard, Fred E.; McKenzie, R. Tait; and Raycroft, Joseph E. *The Aims and Scope of Physical Education.* Archives of the College Physical Education Association.
5. McKenzie, R. Tait. "An Ideal Program of Physical Education in Men's Colleges." Society of Directors of Physical Education in Colleges, *Proceedings,* 1936.
6. ———. *Exercise in Education and Medicine.* Philadelphia: W. B. Saunders Co., 1924.
7. ———. "How Far Should Physical Training Be Educational and How Far Recreative in Colleges and Universities?" *Journal of Proceedings and Addresses of the 44th Annual Meeting of the National Education Association,* 1905.
8. ———. "Physical Education at the University of Pennsylvania—from 1904 to 1931—and the Gates Plan." Society of Directors of Physical Education in Colleges, *Proceedings,* 1931.
9. ———. "Report from the Second District." National Collegiate Athletic Association, *Proceedings* **14**: 11-13, 1919.
10. ———. "The Regulation of Athletics Sports in Colleges." *Proceedings of the American Association for the Advancement of Physical Education,* 1893.
11. ———. "The Relation of Physical and Health Activities to the Academic Program." National Collegiate Association, *Proceedings* **30**: 85-105; 1935.
12. ———. "Variations of Athletic Types." *The Outlook* **131**: 342-43; 1922.
13. Mercer, E. LeRoy. "The Gates Plan." Society of Directors of Physical Education in Colleges, *Proceedings,* 1931.
14. "Presentation of the American Association of Physical Education Fellowship Award." *Journal of Health and Physical Education* **2**: 17; June 1931.

(Submitted 7/18/58)

Self-Attitudes of Women Physical Education Major Students as Related to Measures of Interest and Success[1]

WILMA ISENBERGER
Wisconsin State College
Oshkosh, Wisconsin

Abstract

The purpose of this study was to determine the relationship between the self-attitudes of women physical education major students and measures of interest and success. A second purpose of this study was to determine the relationship between the interests of the women physical education major students and those of women physcial education teachers. Two hundred seventy-seven women physical education major students from three institutions and 115 women physical education teachers participated in the various phases of this study. The "Who Am I?" test was used as a measure of self-attitudes. Interest was measured by the Strong Vocational Interest Blank. Success was measured by the Minnesota Teacher Attitude Inventory, the Scott General Motor Ability Test, a teacher trait evaluation sheet, semester grades in physical education theory and skill courses, and total university semester grades. The results of this study indicated that self-attitudes are not significantly related to interest as measured by the Strong Vocational Interest Blank. Self-attitudes and success as measured by the Minnesota Teacher Attitude Inventory are not significantly related. A negative but significant relationship between judges' ratings of teacher traits and self-attitudes was indicated. The relationship between semester grades and self-attitudes, and between physical education skill grades and self-attitudes was significant for one of the student groups. A relationship between motor ability and self-attitudes was indicated. Women physical education major students scored significantly higher on the Strong Vocational Interest Blank than did the teachers of physical education.

WITHIN THE PAST 20 or 25 years, the area of attitudes has received much attention and consideration. More recently, attitudes toward self have been studied in an attempt to gain a better understanding of individuals and of personality development. Results of such studies reveal that one's self-attitudes are really other-attitudes; that is, an individual tends to see himself as he thinks others see him.

Group and group membership play an important part in attitude development. It seems logical to assume that if the attitudes of a group and of an individual can each be determined, these measurements could be used interchangeably in a predictive manner. Applying this to physical education, if major students as a group possess characteristic attitudes, each individual

---

[1]This report is taken from the findings of a study conducted under the direction of M. Gladys Scott and submitted as partial fulfillment of the requirements for the Doctor of Philosophy degree at the State University of Iowa.

physical education major student would possess similar attitudes. It then becomes evident that an individual wanting to become a part of this group would possess similar attitudes. If it is found that an individual who has entered this group does not display these characteristic attitudes, such a knowledge can be utilized either in guiding the student out of physical education and into another area or in assisting the student toward a better understanding of himself and his profession.

Interest and success cannot be overlooked in the consideration of attitudes. The magnitude of each of these factors would affect the dimensions of attitudes. It therefore becomes essential to consider three elements — interest, success, and self-attitudes—in any attempt to define or describe physical education major students.

This study represents a first attempt to measure and identify the self-attitudes of women physical education major students by employing those techniques used in the social-psychological area. The purpose of this study was to determine the relationship between self-attitudes of women physical education major students and measures of interest and success. A second purpose was to determine the relationship between measures of success of women physical edcation teachers and women physical education major students.

Definition of Terms

Self-attitude as used in this study was based upon Mead's orientation that conceives of the self as being an object and, like all other objects, possessing an attitude or plan of action (15). These attitudes toward self were identified and measured by responses to the Kuhn-McPartland "Who Am I?" test, a Twenty Statements Test (TST). Locus scores, the degree of social anchorage, were used as the numerical representation of self-attitudes.

Interest as defined for this study referred to the special attention to or a concern for physical education teaching as an object. This variable was measured by the Strong Vocational Interest Blank (SVIB).

Success as used in this study referred to specific attainment or achievement and also to a prediction of achievement. This variable was measured by the Minnesota Teacher Attitude Inventory (MTAI), the Scott General Motor Ability Test, first semester grades in physical education theory and skill courses, and total university semester grades. The teacher trait evaluation sheet constructed by Jaeger (11) was also used as a measure of success. It is composed of 12 traits which are considered to be essential characteristics of a good teacher: leadership, interest in learner, sincerity, vitality, adaptability, fairness, dependability, initiative, poise, cooperation, enthusiasm, and sense of humor.

Review of Literature

There is some indication that self-attitudes are related to occupational categories. McPartland states that "occupational expectations may be expected to be involved in present self-attitudes" (14). Stewart (20) obtained evidence in his study which was suggestive of the importance of occupation reference groups in prescribing behavior with respect to one's self. Berg believes that the "actual choice of an occupation is one of the points in life at which a young person is called upon to state rather explicitly his concept of himself and to say definitely, 'I am this or that kind of person'" (1). Burgess (3)

concluded from his study that the degree to which an individual identifies himself with the members of an occupation may be a significant determiner of the degree to which he has the interests of men in that occupation. Hoppock (9), in a study comparing the extreme groups of satisfied and dissatisfied teachers, advanced a similar view. "The determining factor in job satisfaction is most likely to be a synthesis of statuses—the status of the individual within his occupation and the status of the occupation in the community, combining to determine the relative status of the individual in the social economic group with which he identifies himself" (9).

Duggan (7), attempting to determine differences between women physical education major and nonmajor students, obtained results which disclosed consistent differential patterns of interest allied with work, play, and social relationships. Majors differed most from nonmajors in preference for exciting, vigorous, competitive outdoor activities both in work and in play. There was also a difference noted in their reactions to characteristics of people and to the list of prominent women included in a questionnaire. Palmer (16), in diagnosing certain personal qualities of women teachers of physical education and relating the measures to teaching success, concluded that possession of an early interest in teaching physical education did not have any significant degree of relationship to success in teaching. It was also indicated that an early decision to become a teacher of physical education is not an index of success in teaching physical education.

Hurlburt (10), employing the TST in a study of self in relation to a specific role, used 124 nonphysical education major students in bowling classes as subjects. This study seems to be of import to the present investigation. The results which she obtained suggested that the more intensively one becomes involved in an activity, the more it becomes a part of one's self-conception. Significant correlations were also obtained between scores made in bowling and the number of sports in which the subject participated, and between self-attitudes and the extent of immersion of the individual in an area of interest. A relationship was also indicated between the frequency of sports mentioned on the TST and the scores made in bowling over a two-month period.

Blesh (2) obtained results in his study which indicated that an interest in teaching as a profession, intelligence, and academic achievement were attributes which were considered essential to success in teaching physical education. The results from a study by Rostker (18) were similar. Intelligence of teachers, as measured by the specific test employed, is closely related to teaching ability. Attitude towards teaching was found to be significantly associated with teaching ability. Pearson (17), in a study of the comparative performance of undergraduate students in education, obtained results which were similar to those of Rostker. A low positive relationship was indicated between the variables teacher attitude, as measured by the Minnesota Teacher Attitude Inventory, and intelligence.

A survey of the literature has revealed the absence of studies which have used the "Who Am I?" Test (TST) as a specific measure of self-attitudes of women physical education major students and teachers. To this writer's knowledge, there have been no previous attempts to determine the relationship between self-attitudes and the measures of interest and success as used in this study.

Subjects

Two hundred seventy-seven women physical education major students and 115 women physical education teachers served as subjects in this study. One hundred sixty-eight of the participating students were enrolled at Illinois State Normal University, 57 were students at the State University of Iowa, and 52 were students at the University of Nebraska. All of the students did not participate in all phases of this study. In some cases, only data on those classified as junior or senior were collected. In other instances information was not available on each student in departmental files. The teachers participating in this phase of the study were alumnae of either the State University of Iowa or Illinois State Normal University.

Procedure

The TST was administered to the women physical education major students of the three participating schools in the fall of 1955. Immediately following the administration of the TST, the Strong Vocational Interest Blank (SVIB) was administered to the students at the State University of Iowa and at the University of Nebraska. In the spring of 1956 the Minnesota Teacher Attitude Inventory (MTAI) was administered to the junior and senior major students of the University of Nebraska and the junior major students at the State University of Iowa. Staff members of the various institutions administered the test. In an effort to keep the directions uniform, instructions for administration were given to each administrator.

Copies of the TST and the SVIB were mailed to the participating teachers. Instructions for each test were included.

All participating junior and senior students were given a teacher trait evaluation sheet in the spring of 1956. Three instructors familiar with the students and with their experiences also rated each student.

General Motor Ability Scores were obtained for each student from the departmental files in each institution. No attempt was made to administer the motor ability test to those students for whom there were no recorded scores.

The total university semester grade, physical education theory grades, and physical education skill grades for the first semester of 1955-1956 were obtained from department files for the students of the State University of Iowa and of the University of Nebraska. Subjects for whom no grades were available were dropped from the phases of the study involving the grade variable.

Locus scores, the degree of social anchorage or the degree of importance of things external to oneself, were computed from the responses to the TST. The method of scoring developed by Kuhn and McPartland was used (12). This procedure necessitated interpreting each of the responses on each questionnaire as either *consensual*, i.e., reference to group membership such as girl, student, major in physical education; or *subconsensual*, i.e., references to groups or feelings whose limits could be identified by only the respondent. Examples of the latter type are a good student, lover of sports, happy individual. The point in the 20 responses at which the respondent ceased giving consensual statements was the numerical representation of social anchorage, the expression of dependence upon things or groups outside himself.

The mean, standard error of the mean, and standard deviation of the locus scores for each of the groups are reported in another study.

The SVIB was machine scored. Only the scoring scale for high school teachers of physical education as developed by Collins (5) was used. All answers were in terms of standard scores. The mean, standard error of mean, and standard deviation were computed for each of the groups. The "t" test was used to determine the significance of the difference between the means.

The MTAI was scored by hand. The final score was the number of correct responses minus the incorrect. T-scores for each subject were obtained from the appropriate scales constructed by Cook (6). Mean, standard error of the mean, and standard deviation were computed for the distribution. The "t" test was used to indicate the significance of the difference between the means.

General Motor Ability Scores, as used, were in terms of T-scores on the particular battery of tests administered at each institution. The State University of Iowa (School B) and the University of Nebraska (School C) used a three-item battery composed of basketball throw for distance, obstacle run, and the standing broad jump. Illinois State Normal University (School A) used the four-item battery composed of the basketball wall pass, basketball throw for distance, four-second dash, and the standing broad jump. The total scores for each battery were computed according to the regression equations presented by Scott and French (19).

The scores for the 12 items on the teacher trait evaluation sheet were totaled for each of the self-evaluations and for the judges' evaluation. An average of the judges' evaluation was computed.

The Pearson Product Moment Method of Correlation (8) was used to determine the relationship between the variables of self-attitudes, interest, and success. Scores for each school on most variables were kept separate.

Analysis of Data

The mean scores obtained by the various groups on the SVIB are presented in Table 1. Reference to the "t's" of Table 2 indicate no significant difference between the student groups within a school, between schools, and between the various teacher groups. A significant difference was found between

TABLE 1.—SUMMARY OF MEAN SCORES ON SVIB OF STUDENT AND TEACHER GROUPS

School	N	M	S.E.	S.D.
Senior				
B	14	42.64	1.80	5.51
C	12	41.33	2.01	6.67
All	26	42.04	1.33	6.63
Junior				
B	10	39.50	2.35	7.06
C	10	40.10	2.59	7.78
All	20	39.80	1.71	7.43
Sophomore				
B	15	42.47	1.73	6.46
C	11	41.18	1.40	4.43
All	26	41.92	1.15	5.73
Freshman				
B	18	41.00	1.70	7.03
C	11	41.18	1.48	4.67
All	29	41.07	1.23	6.51
Total				
B	57	41.53	.91	6.84
C	44	40.98	.95	6.22
All	101	41.29	.66	6.60
Teachers				
Group				
I	7	41.43	2.86	7.01
II	19	38.11	1.26	5.35
III	31	38.39	1.08	5.90
IV	58	35.52	1.04	7.86
All	115	37.08	.67	7.17

the means of the total teacher group and the senior, sophomore, freshman, and total groups of both School B and School C. A significant difference was also obtained between these same groups of total students and total teachers. These results seem to indicate that the scale as now used either does not give an accurate picture of the interests of the physical education teacher group or that similarity of interests decrease as one goes into the teaching field.

The scoring scale used was constructed for high school teachers of physical education. Included in the teacher group serving as subjects in this study were elementary, high school, junior college, and college teachers of physical education, as well as elementary and high school supervisors and administrators at the high school and college level. Each of these groups might possess interests which would differentiate it from the other as well as from women in other occupations. If such is the case, it would appear that students also would express different interests, depending upon their future teaching plans. Another factor to be considered is that the interests of women in physical

TABLE 2.—COMPARISON OF MEAN SCORES ON SVIB OF STUDENTS ACCORDING TO CLASSIFICATION AND SCHOOL, AND WITH TEACHER GROUPS

Group	Senior	Junior	Sophomore	Freshman	Total
School B					
Junior	1.06				
Sophomore	.07	1.02			
Freshman	.66	.52	.61		
Total	.55	.80	.48	.27	
School C	.48	.17	.58	.08	.42
Total Students	.27	.10	.27	.03	.21
Total Teachers	2.89[a]	.99	2.91[a]	2.14[b]	3.92[a]
School C					
Junior	.37				
Sophomore	.62	.37			
Freshman	.60	.36	2.03		
Total	1.57	.32	.12	.11	
Total Students	.29	.10	.41	.06	.27
Total Teachers	2.00[b]	1.13	2.64[a]	2.53[b]	3.36[a]
Total Teachers and Total Students	3.34[a]	1.48	3.65[a]	2.85[a]	4.47[a]
Teachers					
	Group I	Group II	Group III	Group IV	
Group II	1.06				
Group III	.99	.17			
Group IV	1.94	1.59	1.92		
Total	1.48	.72	1.03	1.26	

[a] Significant at 1 percent level of confidence.
[b] Significant at 5 percent level of confidence.

education could have changed since the publication of the scoring key. A general change of attitude toward woman and her place in society has occurred since the 1940's. Further research needs to be done to determine the present appropriateness of the scoring key.

Strong obtained results from his study which indicated that there was an increase in mean scores within the first five years after leaving college, and that any changes which did occur after this interval were in a downward direction (21). There is an indication that this is true with the teacher groups as used in this study. However, the difference in the size of the groups at each level of experience does not warrant either the acceptance or rejection of this assumption.

It is not possible to explain why the junior groups obtained lower mean scores than the other three classification groups. There may be some element present at this level which would account for this lower score. The group may have attempted to fake their scores. Comments received from the junior group at one institution revealed that they had intentionally responded in a manner which they believed to be opposite the expected pattern. Strong found that it is possible to fake scores on the SVIB, but such faking is usually

TABLE 3.—SUMMARY AND COMPARISON OF MEAN SCORES ON MTAI OF STUDENT GROUPS

Group	N	M	S.E.	S.D.	"t"
School B	12	56.08	7.29	24.21	
School C	18	40.79	8.09	33.36	1.07
Seniors	11	39.09	9.24	29.22	
Juniors	19	64.37	6.23	26.42	2.27[a]

[a] Significant at 5 percent level of confidence.

TABLE 4.—RELATIONSHIP BETWEEN MEASURES OF SELF-ATTITUDES, INTEREST, AND SUCCESS

School A

Variable[a]	I	II	VI
VI	—.06	—.07	
VII	—.01	.06	—.03
VIII	—.17	—.41	.33
Subject Numbers for Each Variable			
VI	121	98	
VII	45	36	45
VIII	42	36	34

[a] I —Locus score, first administration.
II —Locus score, second administration.
VI —Motor ability.
VII —Teacher trait evaluation, judges' ratings.
VIII—Teacher trait evaluation, self ratings.

School B

Variable[a]	I	IX	VII	VIII	X
IX	.19				
VII	—.02	.14			
VIII	—.15	.12	.44		
XI	.28[c]	.07	.61[b]	.20	.18
XII	.15	.12	.58[b]	.21	.02
XIII	.45[b]	.20	.42	—.21	.04
VI	.23	.01	.11	.04	
Subject Numbers for Each Variable					
IX	57				
VII	20	20			
VIII	21	21	19		
XI	52	52	20	20	11
XII	35	35	20	20	11
XIII	50	50	19	19	11
VI	52	52	17	19	

[a] I —Locus score, first administration.
VI —Motor ability.
VII —Teacher trait evaluation, judges' ratings.
VIII—Teacher trait evaluation, self ratings.
IX —SVBI.
X —MTAI.
XI —Semester grades, 1955-1956.
XII —Semester grades in physical education theory, 1955.
XIII—Semester grades in physical education skill, 1955.

[b] Significant at 1 percent level of confidence.
[c] Significant at 5 percent level of confidence.

School C

Variable[a]	I	VII	VIII	X
VI	.02	−.40	.14	
VII	−.22			
VIII	−.11	.02		
IX	.22			
XI	−.08	.51[c]	.18	−.23
XII	.29	.44	−.06	.31
XIII	−.09	.61[a]	.18	−.03
Subject Numbers for Each Variable				
VI	33	18	17	
VII	17			
VIII	17	21		
IX	40			
XI	34	20	19	15
XII	33	20	19	15
XIII	33	19	18	14

[a] I —Locus score, first administration.
VI —Motor ability.
VII —Teacher trait evaluation, judges' ratings.
VIII—Teacher trait evaluation, self ratings.
IX —SVIB.
X —MTAI.
XI —Semester grades, 1955-1956.
XII —Semester grades in physical education theory, 1955.
XIII—Semester grades in physical education skill, 1955.

[b] Significant at 1 percent level of confidence.
[c] Significant at 5 percent level of confidence.

TABLE 5.—RELATIONSHIP BETWEEN MTAI AND MEASURES OF SELF-ATTITUDE, INTEREST, AND SUCCESS

Schools B and C Combined

Variable[a]	I	VI	VII	VIII	IX
MTAI	−.08	−.21	.29	−.10	−.22
N	26	27	27	29	26

[a] I —Locus score, first administration.
VI —Motor ability.
VII —Teacher trait evaluation, judges' ratings.
VIII—Teacher trait evaluation, self ratings.
IX —Strong's vocational interest blank.

TABLE 6.—RELATIONSHIP BETWEEN MOTOR ABILITY AND LOCUS SCORE

Total Students
N 206

Variable	Locus
Motor Ability	.11

TABLE 7.—RELATIONSHIP OF LOCUS SCORES AND YEARS OF EXPERIENCE, AND LOCUS SCORES AND SVIB SCORES OF TEACHERS

Variable	SVIB	Years of Experience
Locus	.14	.06
N	115	115

in an upward direction (21). Longstaff also found this to be true (13).

The scores obtained on the MTAI indicate a significant difference between the junior and senior groups. (See Table 3.) Why the junior group should attain a mean score higher than that of the senior group cannot be explained. Group size may be a factor. A larger number would eliminate some of the extremes which are over-effective in small samples. Callis (4), as a result of his study using the MTAI, concluded that teacher-pupil attitudes are operative in the subject's selection of the field of education in which he wishes to specialize, and it is possible that these attitudes have elements in common with vocational interest. The results obtained in this study would not support this assumption. The scores attained by the juniors on the MTAI were significantly higher than those of the seniors. On the SVIB, the reverse was true, although not significantly so. The relationship between scores obtained on both the MTAI and the SVIB was not significant but was in the negative direction. (Refer to Table 5.)

The correlation coefficients between the different variables are presented in Tables 4-7. The corresponding number of subjects for whom data were available for correlation purposes is listed in a separate table. Group sizes in this particular phase of the study prevent a true picture of relationship.

A low but significant correlation was evident between the variables locus score and semester grades and between the locus score and physical education skill grades of the School B subjects. The obtained correlation coefficients for these variables for the School C subjects indicate a slight inverse relationship. It appears that the relationship between locus scores and success in school work as measured by grades is greater in a liberal arts college where a broader educational background is emphasized.

A significant relationship was found between judges' ratings of teacher traits and of semester grades for both School B and School C. Similar relationship was found to exist between judges' ratings of teacher traits and physical education skill grades at School C and between judges' ratings and physical education theory grades at School B. Grades, then, play an important part in the interpretation of teacher traits.

Conclusions

The following conclusions are based upon the data obtained in this study:

1. Self-attitudes, represented in terms of locus scores, are not significantly related to interest.

2. The relationship between self-attitudes and success, as measured by the Minnesota Teacher Attitude Inventory, is not significant.

3. Judges' ratings of teacher traits, as another measure of success, is negatively related to self-attitudes. However, the obtained correlation coefficients are not of a dimension to be significant.

4. There is a significant relationship between semester grades and self-attitudes of one student group. The relationship between these variables is

not significant and tends toward the negative for the other group studied.

5. Physical education theory grades tend to be positively related to self-attitudes.

6. Grades in physical education skills and self-attitudes of one student group are significantly related. However, for the second group studied on this variable, the relationship tends to be of negative quality.

7. There is a small but positive relationship between motor ability and self-attitudes.

8. The women physical education major students scored significantly higher on the interest test than did the women teachers of physical education.

References

1. Berg, Irwin. "Personality Structure and Occupation Choice." *Personnel Guidance Journal* **32**: 151-54; 1953.
2. Blesh, T. E. "Correlations between Success in Student Teaching and Success on the Job." *Research Quarterly* **13**: 397-99; Oct. 1942.
3. Burgess, Thomas C. *A Study of Certain Relationships Between Self-Concept Vocational Interests, and Occupational Stereotypes.* Unpublished Doctoral dissertation. Minneapolis: University of Minnesota, 1954.
4. Callis, Robert. "Change in Teacher-Pupil Attitudes Related to Training and Experience." *Educational and Psychological Measurement* **10**: 718-27; 1950.
5. Collins, Patricia J. "The Development of a Scoring Key on the Strong Vocational Interest Inventory for Women Teachers of Physical Education." *Research Quarterly* **13**: 156; May 1942.
6. Cook, Walter W., and Callis, Robert. *Minnesota Teacher Attitude Inventory Manual.* New York: Psychological Corp., 1951.
7. Duggan, Anne Schley. "A Comparative Study of Undergraduate Women Majors and Non-Majors in Physical Education With Respect To Certain Personal Traits." *Research Quarterly* **8**: 38-45; Oct. 1937.
8. Guilford, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill Book Co., 1950.
9. Hoppock, Robert. *Job Satisfaction.* New York: Harper and Brothers, 1935.
10. Hurlburt, Julia Knaff. *An Empirical Study of the Self in Relation to a Specific Role.* Unpublished Master's thesis. Iowa City: State University of Iowa, 1956.
11. Jaeger, Eloise M. *An Investigation of a Projective Test in Determining Attitudes of Prospective Teachers of Physical Education.* Unpublished Doctoral dissertation. Iowa City: State University of Iowa, 1952.
12. Kuhn, Manford, and McPartland, Thomas S. "An Empirical Investigation of Self-Attitudes." *American Sociological Review* **19**: 68-76; 1954.
13. Longstaff, Howard P. "Fakability of the Strong Vocational Interest Blank and Kuder Preference Record." *Journal of Applied Psychology* **32**: 360-69; 1948.
14. McPartland, Thomas S. *The Self and Social Structure: An Empirical Approach.* Unpublished Doctoral dissertation. Iowa City: State University of Iowa, 1953.
15. Mead, George H. *Mind, Self and Society.* Chicago: University of Chicago Press, 1934.
16. Palmer, Irene. "Personal Qualities of Women Teachers of Physical Education." *Research Quarterly* **4**: 31-47; Dec. 1933.
17. Pearson, Avid Neil. *A Study of the Relationships Between Attitudes, Intelligence. and Achievement of University of Minnesota Majors in Education Grouped by Major Area of Specialization.* Unpublished Doctoral dissertation. Minneapolis: University of Minnesota, 1955.

18. Rostker, Leon E. "The Measurement and Prediction of Teaching Ability." *School and Society* **51**: 30-32; 1940.
19. Scott, M. Gladys, and French, Esther. *Evaluation in Physical Education.* St. Louis: C. V. Mosby Co., 1950.
20. Stewart, Robert L. *An Empirical Test of the Reference Group Hypothesis.* Unpublished Master's thesis. Iowa City: State University of Iowa, 1953.
21. Strong, Edward K. *Vocational Interests of Men and Women.* Stanford, California: Stanford University Press, 1943.

(Submitted 8/14/58)

Substitution of the Tensiometer for the Dynamometer in Back and Leg Lift Testing[1]

FRANK T. KENNEDY
Point Grey Junior High School
Vancouver, British Columbia

Abstract

This study investigates the possibility of substituting the relatively inexpensive cable tensiometer for the back and leg dynamometer in the Rogers Physical Fitness Index Test. A lever system was devised to overcome the insufficient capacity of the tensiometer for the leg lift. Both the lever system and the tensiometer were calibrated against the criterion of the dynamometer. College men performed with both instruments on a test-retest pattern to determine by correlational means if the two tests were in agreement, and the means and standard deviations obtained by the subjects on the two instruments were compared. Analysis of the results indicated that the tensiometer may be substituted for the dynamometer in back and leg lift testing.

The Problem

The dynamometer might well be termed the cornerstone of strength testing in physical education. Following the invention of this instrument by Regnier in France (1), Sargent (8, 9, 10) developed his Intercollegiate Strength Test at Harvard University in 1880. Nearly 50 years later, Rogers (7) revised this test and, in doing so, constructed norms for the strength index (SI), thus creating the Physical Fitness Index (PFI). Some revisions have been made in the original Rogers version of the test, although the test items remain the same. Two essential tests in both the Sargent and the Rogers batteries are the back and leg lifts performed with a dynamometer.

During the past decade, several other instruments have been introduced in the field of strength testing. Among these are the cable tensiometer, the strain gauge, the load cell, and the myometer. Certain of these instruments have advantages over the dynamometer; they are small and easy to operate, and, most of all, they are much less expensive. A consistent drawback in the general use of the Rogers PFI test has been the cost of the back and leg dynamometer. This study, therefore, was undertaken to determine if the cable tensiometer could be substituted for the dynamometer in back and leg lift

[1]This study was made as a Master's thesis in the School of Health and Physical Education at the University of Oregon. The writer wishes to express his sincere appreciation to H. Harrison Clarke for his assistance and guidance in writing both the thesis and this article, and to L. R. Geser and G. Carter for their invaluable assistance in the administration of the tests.

testing without appreciable change in test scores achieved by college men on these two tests.

Testing Instruments

Dynamometer. The back and leg dynamometer consists of two elliptical springs joined at their extremities. Between these springs is a large dial calibrated in pounds. Two hands on the face of the dial indicate the amount of deformation in the shape of the springs when a lift is made. One of the hands records the point of maximum lift and must be manually returned to the zero position after each lift.

When using this instrument for back and leg strength tests, one end of the springs is fixed to the bench and the other is attached to a chain through which the force of the lift causes a stretching action of the springs. The dynamometer selected for this study had a maximum capacity of 2500 lb.

Tensiometer. The cable tensiometer was originally designed to measure the tension of aircraft control cable. Cable tension is determined by measuring the amount of force needed to create offset in the cable by the two sectors and a riser. The amount of offset is recorded on the dial of the tensiometer and is converted directly to pounds by a calibration chart.

This tensiometer was adapted to strength testing by Clarke and Peterson in 1945 (2). The purpose in the initial use of the instrument was to test the strength of muscle groups involved in orthopedic disabilities. Clarke and associates (3) have developed 38 cable-tension strength tests involving most of the muscle groups of the body (5). Subsequently, however, the tensiometer has been utilized in testing the strength of individual muscle groups without reference to disabilities.

The tensiometer used in this study had a maximum capacity of 800 lb.; a 3/32-in. aircraft cable with a tensile strength of 1800 lb. was used. The capacity of this instrument was quite sufficient for the back lift, but cases of leg lifts up to 2400 lb. are not unknown when using the belt technique. Since tensiometers are not manufactured with this capacity and in order to utilize the one tensiometer for both lifts, a lever system was designed for use in the measurement of the leg lift.

Leg Lift Test with Tensiometer

Lever System. Several problems were encountered in the construction of a lever system to permit the use of a tensiometer with a maximum capacity of 800 lb. in the administration of leg lifts which conceivably could reach 2400 lb. The procedures were as follows:

1. The primary concern was to find material for the lever arm capable of withstanding a stress of 2400 lb. After some investigation, Shelby steel tubing, 1 in. in diameter and 12 in. long, was selected as the most suitable material because of its great strength and relatively low cost.

2. Since there was some question of whether the Shelby tubing could be drilled for the fulcrums and still withstand the same stress, movable collars

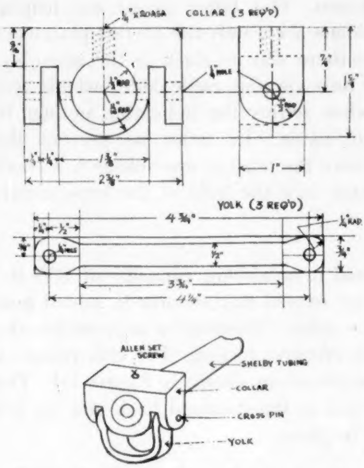


FIGURE I. Specifications for construction of collars.

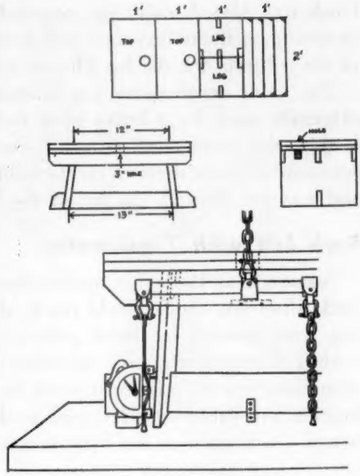


FIGURE II. Leg lift bench for lever assembly.

permitting each of the three fulcrums to remain in the centre of the long axis of the tubing were utilized. These collars were constructed of chrome-molybdenum steel with Allen wrench setscrews to secure their placement. Specifications for the construction of the collars are contained in Figure I.

3. In arranging the lever system, two-thirds of the lift was to be taken by a chain strong enough to withstand stresses of at least 2000 lb. and one-third of the lift would be applied to a 3/32-in. cable upon which would be placed the tensiometer. With a 12-in. lever arm, the distance between the centers of the two end fulcrums was 11 in. Applying physical principles, the center of the inside fulcrum was located $3\frac{2}{3}$ in. from the center of the chain fulcrum and $7\frac{1}{3}$ in. from the center of the cable fulcrum. The arrangement of the lever system may be seen in Figure II.

Bench Construction. During the early trials of the lever assembly in the administration of the leg lift, testers found that they could not keep the tensiometer in place while raising the lever arm and hooking the chain to the lifting bar; and the estimation of the correct placement of the subject was difficult. For these reasons, a bench was constructed from $\frac{3}{4}$-in. plywood under which the lever assembly was suspended, as shown in Figure II. All joints were fastened with glue and screws and attached to a base, 24" x 36" x $1\frac{1}{2}$", by small angle iron brackets and screws. Holes were drilled through the base to secure the fixed chain and the cable, and a steel cross pin prevented these two anchors from pulling out of the base.

To maintain a slight tension on the assembly when not in use, a supporting

block was placed under the centre fulcrum. This latter device was helpful in testing, as the testers were able to estimate accurately the correct placement of the subjects for the leg lift test since there was no slack in the assembly.

The cable tensiometer has a small hole on the cable side, which was originally used for a brake lever rod when setting the indicating pointer in testing the tension of aircraft control cable. In order to prevent the tensiometer from slipping on the cable once the tension was released, a small rod was put through the leg of the bench into the hole of the tensiometer.

Back Lift with Tensiometer

Inasmuch as the cable tensiometer had a maximum capacity of 800 lb., back lifts tests, which would rarely if ever exceed this amount in school testing, were possible by direct pulls on the cable. However, a support for the cable and tensiometer was necessary for effective testing. For this reason a strap iron bracket was constructed to specifications shown in Figure III. The bracket and cable were fastened to the rod in the standard back and leg lift bench which anchors the dynamometer in place.

Calibration of Equipment

To be certain that all instruments and devices would produce identical scores for the same application of force, a common calibration was provided. A mechanical device for creating forces from 100 to 2400 lb. was necessary for the full range of strength scores anticipated in this study. A movable engine hoist, utilized in garages for removing engines from automobiles, was chosen for this purpose. Two separate calibrations were made: the first with

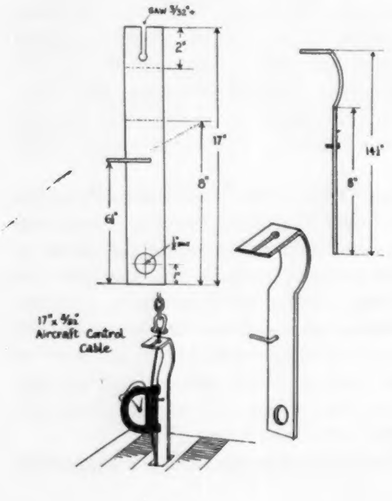


FIGURE III. Bracket for tensiometer in back lift.

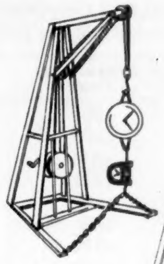


FIGURE IVa. Calibration of tensiometer.

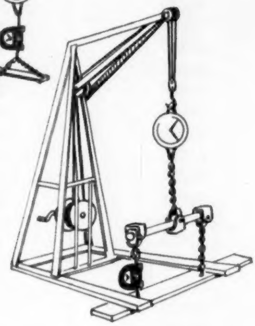


FIGURE IVb. Calibration of lever assembly.

dynamometer and tensiometer, and the second with the dynamometer and the lever assembly.

Tensiometer Calibration. The calibration of the tensiometer was accomplished against the back and leg dynamometer as shown in Figure IVa. In this process, the dynamometer was hung from a hook on the winch with the aircraft cable attached beneath and fastened by a chain to the base. The tensiometer was placed on the cable and the winch was tightened by cranking; readings on the tensiometer were recorded at 50-lb. intervals, as indicated by the dynamometer until 800 lb. was reached. After each reading, the tension was released and the maximum pointers were returned to zero in preparation for the next stress. In this way, stresses and recordings were made under simulated "lifting" conditions, and damage to either instrument by sustained stress was prevented. As a check on the accuracy of the process, the calibration was repeated three times. In each case, the results were identical with the first reading and therefore assumed to be correct. For easy conversion of the tensiometer readings to pounds, Table 1 was constructed.

Lever Assembly Calibration. The lever assembly was calibrated in a manner similar to the above, but at intervals of 100 lb. from 600 to 2400 lb. as illustrated in Figure IVb. Because of the extreme stress on the hoist frame required for the upper range of the calibration, only two complete trials were made. The tensiometer readings were recorded in pounds read from the dynamometer to enable the lever-assembly readings to be converted directly into pounds pull (as the actual stress is three times the recorded tensiometer reading). The two readings taken for the lever assembly were in agreement over the entire range. Table 2 was prepared for easy conversion of the tensiometer readings.

TABLE 1.—DIRECT CONVERSION CHART FOR BACK LIFT TENSIOMETER TO DYNAMOMETER-POUNDS

Tensiometer Reading	Dynamometer Pounds	Tensiometer Reading	Dynamometer Pounds
17	100	54	360
18	105	55	370
19	110	56	380
20	115	57	390
21	120	58	400
22	125	59	405
23	135	60	420
24	140	61	430
25	145	62	435
26	150	63	450
27	160	64	460
28	165	65	470
29	170	66	480
30	180	67	490
31	185	68	500
32	190	69	510
33	200	70	520
34	205	71	530
35	210	72	550
36	220	73	560
37	225	74	570
38	230	75	580
39	240	76	600
40	250	77	610
41	255	78	625
42	265	79	640
43	270	80	650
44	280	81	670
45	290	82	685
46	300	83	700
47	305	84	715
48	310	85	730
49	320	86	750
50	330	87	760
51	340	88	770
52	350	89	790
53	355	90	800

Experimental Procedures

SUBJECTS

The subjects were nondisabled male students at the University of Oregon. They were selected at random from the nonprofessional physical education classes.

TABLE 2.—DIRECT CONVERSION CHART FOR LEG LIFT LEVER-ASSEMBLY TO DYNAMOMETER-POUNDS

Lever Assembly Reading	Dynamometer Pounds	Lever Assembly Reading	Dynamometer Pounds
33	600	64	1315
34	620	65	1350
35	640	66	1375
36	660	67	1400
37	680	68	1425
38	700	69	1450
39	725	70	1485
40	750	71	1515
41	775	72	1565
42	800	73	1600
43	820	74	1635
44	840	75	1675
45	860	76	1725
46	880	77	1750
47	900	78	1800
48	925	79	1825
49	950	80	1875
50	975	81	1910
51	1000	82	1975
52	1025	83	2010
53	1050	84	2060
54	1065	85	2100
55	1085	86	2150
56	1110	87	2200
57	1130	88	2225
58	1145	89	2275
59	1175	90	2325
60	1200	91	2365
61	1225	92	2400
62	1255	93	2450
63	1285	94	2500

TESTERS

The testers used in this study were assistants in the Physical Education Research Laboratory. All testers had had extensive training and experience in administering back and leg lift tests with the dynamometer. As will be seen below, the agreement between testers on repeated tests was found to be satisfactory.

TEST DESCRIPTION

The back and leg lift testing techniques were in accordance with the standard procedures described by Clarke (4). Briefly stated, the positions were as follows:

Back Lift. The subject lifted from a standing position with the back slightly bent and the knees straight.

Leg Lift. For this test, the knees were slightly bent and the back was straight, the head erect and the chest up. A belt was used to clamp the lifting bar to the subject's body.

OBJECTIVITY

Objectivity coefficients for the back and leg lift tests with the tensiometer, utilizing the special devices and assemblies described herein, were obtained in the following manner: two testers administered each of these tests independently to 30 subjects, and the coefficient of correlation was computed to determine the degree of tester agreement for each test.

COMPARISON OF INSTRUMENTS

After the objectivity coefficients for the new test devices were determined, the investigator administered back and leg lift tests to 36 subjects with both the dynamometer and the tensiometer arrangement. To determine the comparability of these two testing procedures, the following statistical analyses were employed for each of the tests, the back lift and the leg lift:

1. Coefficients of correlation were computed between the scores obtained with the two instruments in order to show the degree of general agreement throughout the range of scores.
2. The mean score obtained with each instrument was determined, and the difference between means was tested for significance.
3. Similarly, the standard deviations for each instrument were computed and the difference between standard deviations was tested for significance.

The significance of the differences between the means and standard deviations was tested by the t ratio. As this study was in the nature of a one-group experiment, the groups were considered correlated. The degrees of freedom were $N - 1$, or $36 - 1 = 35$. For degrees of freedom, a t of 2.03 is significant at the .05 level (6). In this study, the .05 level was chosen as indicating an undesirable difference between means and standard deviations.

Results

BACK LIFT

Objectivity. The objectivity coefficient obtained when the two testers administered the back lift with the tensiometer was .90. This coefficient compares favorably with the test-retest reliability estimate, .88, reported for back lift testing with the dynamometer (5).

Correlation between Instruments. The coefficient of correlation between back lift scores when the investigator tested the same subjects with the tensiometer and with the dynamometer was .92. Thus, the relative agreement in results obtained for this test with the two instruments is indicated.

Comparison of Means and Standard Deviations. The means and standard deviations obtained from testing the subjects with both the tensiometer and dynamometer were as follows:

	Tensiometer	Dynamometer	Diff.	S. E. Diff.	t Ratio
Mean	525.34	509.50	15.84	11.79	1.34
Standard Deviation	180.74	165.84	14.90	11.31	1.32

The tensiometer yielded slightly higher back lift scores and slightly greater dispersion of scores. However, these differences are not statistically significant at the .05 level of confidence. Thus, it may be concluded with some assurance that comparable results are obtained with the tensiometer and dynamometer for the back lift test.

LEG LIFT

Objectivity. The objectivity coefficient, representing agreement between testers for the leg lift utilizing the lever assembly and tensiometer, was .84. This coefficient is slightly lower than the reliability estimate, .86, reported for leg lift testing with the dynamometer (5). However, some difficulty with the maximum pointer on the tensiometer was experienced; this was corrected before additional testing was done with the instrument.

Correlation between Instruments. The investigator obtained a .95 coefficient of correlation between leg lift scores for the same subjects with the tensiometer and the dynamometer. Thus, the relative agreement in results obtained for this test with the two instruments is high.

Comparison of Means and Standard Deviations. The means and standard deviations obtained from testing the same subjects with the tensiometer and the dynamometer were as follows:

	Lever Assembly	Dynamometer	Diff.	S. E. Diff.	t Ratio
Mean	1143.94	1146.72	2.78	12.04	.23
Standard Deviation	230.87	216.65	14.22	11.579	1.23

The tensiometer yielded nearly identical means for the leg lift scores; the t was .23, indicating a negligible difference between means. The tensiometer produced a greater dispersion than the dynamometer; however, the t for the difference of 14.22 pounds between the standard deviations was 1.23, which is not significant at the .05 level.

Discussion

Both the testers and the subjects preferred the cable tension method in administering the back and leg lift tests. Since this method permits very little excursion (i.e., stretch or displacement) when lifting, the subject remained at his maximum lifting angle throughout the entire lift. In comparison, the back and leg dynamometer allows relatively greater excursion, with the result that the subject may lift past his maximum efficiency angle. The relative durabilities of the tensiometer and the dynamometer were not studied. How-

ever, it is judged that the dynamometer set-up would stand up better under prolonged use. The analysis of the results indicated that the tensiometer may be substituted for the dynamometer in back and leg lift testing.

References

1. Clarke, H. Harrison. *Application of Measurement to Health and Physical Education.* 2nd ed. New York: Prentice-Hall, 1950.
2. ———. "Comparison of Instruments for Recording Muscle Strength." *Research Quarterly* **25**: 398-411; Dec. 1954.
3. ———. *A Manual: Cable Tension Strength Tests.* Chicopee, Mass.: Brown-Murphy Co., 1952.
4. ———. *Research Projects Utilizing Objective Strength Tests by Cable Tension Method.* (Quotation from Binet and Vaschide.) Unpublished Office of Naval Research Report, 1950.
5. ———; Bailey, T. L.; and Shay, C. T. "New Objective Strength Tests of Muscle Groups by Cable Tension Methods." *Research Quarterly* **23**: 136-48; May 1952.
6. Garrett, Henry E. *Statistics in Psychology and Education.* 4th ed. New York: Longmans, Green and Co., 1953.
7. Rogers, Frederick Rand. "A Review of Recent Strength Testing Literature." *Journal of Health and Physical Education* **5**: 8; Mar. 1937.
8. Sargent, Dudley A. "Intercollegiate Strength Tests." *American Physical Education Review* **2**: 216; 1897.
9. ———. "Twenty Years of Progress in Efficiency Tests." *American Physical Education Review* **17**: 453; 1913.
10. Seaver, Jay W. *Anthropometry and Physical Examination.* New Haven: Curtiss-Way Co., 1909.

(Submitted 9/10/58)

The Significance of Interpolated Time Intervals on Motor Learning[1]

M. DOROTHY MASSEY
University of Rhode Island

Abstract

This study was an exploration of the significance of the interpolation of three different time intervals between practices in learning a motor skill. The stabilimeter was the instrument used. Three equated groups followed one of three time patterns for five weeks and had a retention test two weeks later. The subjects were housed in convents, so unique experimental controls were possible. Two of the time patterns simulated those frequently used in educational scheduling. The third time pattern was established by using a "Summation-Series." Significant differences appeared in performance at various times during the experiment.

Introduction

The phenomena of learning and retention have long attracted the interest of research workers. In our society the speed at which a person learns and the length of time he retains the material learned often become the determining criteria for his position in the culture and, in some cases, for his very survival.

Efficiency in the learning process has become a topic of paramount significance in our increasingly competitive world. Although the armed forces, industry, and business are greatly concerned with efficiency in learning, it would seem that the educator should be more concerned because the fundamental problem of learning is the very core of the educator's world.

In education the increasing enrollment and the shortage of teachers result in an accentuation of the demand for more efficient teaching methods. It appears that further research on more efficient learning techniques not only would be helpful to society in general but also would be particularly beneficial to the organized educational world of today and of the future.

In physical education the need for more efficient teaching methods, including the use of the most beneficial time patterns in scheduling, became apparent to some many years ago. Among the reasons for this is the fact that, in this area of education, large classes have been the rule rather than the exception. The pupil-teacher ratio has been traditionally high and will undoubtedly become even higher. In addition, there seems to exist a perpetual lack of well-educated teachers of physical education. The program is one in which motor skills predominate so that the results are often very obvious. It appears that

---

[1]This research was conducted under the direction of John M. Harmon, Boston University, in partial fulfillment of the requirements for the degree of Doctor of Education, 1957.

physical education is a phase of education that is in need of and greatly interested in the outcomes of more research on time patterns in learning.

Review of Literature

For some time the temporal relationships involved in the learning process have been recognized as being important with regard to the speed of learning and the amount of retention. There has been considerable research involving the manipulation of the time interval. The majority of the studies have found that some form of "distributed practice" is generally more efficient than "massed practice."

The results of several previous investigations of the effects of interpolating various lengths of time between practices are of particular significance to this study. Lashley (2), Murphy (5), Miller (4), and Young (8) worked with actual motor skills that might be employed in a physical education program. One could conclude from their results that activities are not always being scheduled in the most efficient and economical manner.

Lashley (2) found in archery target shooting that there was improvement during the first half of the practices regardless of the distribution of time. During the second half of the practices he found greater improvement with greater distribution of the trials. Murphy (5) found that in learning the javelin throw a schedule of practice three days a week had a higher potential for success than a schedule in which practice was held on five days of the week.

Later, Miller (4) reported on his research in which billiards was the skill and four different time patterns were employed. He concluded that the most promising schedule seemed to be one with relative massing at the beginning. Then, after a foundation had been laid, greater spacing between practice periods appeared more favorable than continued massed practice.

Snoddy (6), who was particularly interested in the influence of the time interval on the learning process, conducted a series of experiments with the stabilimeter and presented his analysis of the learning curve which he divided into two parts: an "adaptation" portion and a "facilitation" part which begins at the end of the adaptation. Snoddy described the improvement during the adaptation phase as a positive function of the length of the interval and the improvement during the facilitation stage as a function of the number of repetitions. He interpreted improvement during adaptation as being dependent upon growth processes leading to the formation of a neuromuscular pattern, and facilitation as growing out of the condensation of this pattern.

Doré and Hilgard (1) concluded that the growth between practices was due to the stimulation that occurred during the practice periods and not to aging. They also suggested that the optimum distribution of practice must lie somewhere between the overcrowding which disrupts practice and that separation which allows a loss of previous gain before practice is resumed.

Procedure

At the outset it was decided that a laboratory type of motor skill, which would lend itself readily to the experimental design, should be employed. In addition, the testing instrument should be one that had been proved both valid and reliable. The stabilimeter, which was used extensively by both Snoddy (7) and Lorge (3), fulfilled both these requirements. Figure Ia,b is a drawing of the instrument that was used in the study.

Since the majority of the previous work had involved subjects within the college age group, it appeared that the results would lend themselves more readily to comparisons if subjects of a similar age were again utilized. In order to work with the desired age group, and in order to assure the controls with regard to the type of environment, diet, rest, health, and schedule, girls who were training to be teaching nuns were selected as subjects for this study.

The study was conducted in two convents within a 25-mile radius of Boston, Massachusetts. Both were convents of teaching orders of sisters of the same religious denomination. The subjects used in the study were novice nuns whose ages ranged from 17 to 21. The type of life, environment, educational program, diet, and the amount of rest were quite similar for all the subjects throughout the study. They appeared to be in an excellent state of health. In assembling the groups the entire population was used.

This study was an exploration of the effect of three different interpolated time patterns on the learning of a motor skill. Two of the time patterns tested were of the types typically and currently used in education. The third pattern was the summation series which is frequently referred to in the natural sciences. The retention was checked for all subjects two weeks after comple-

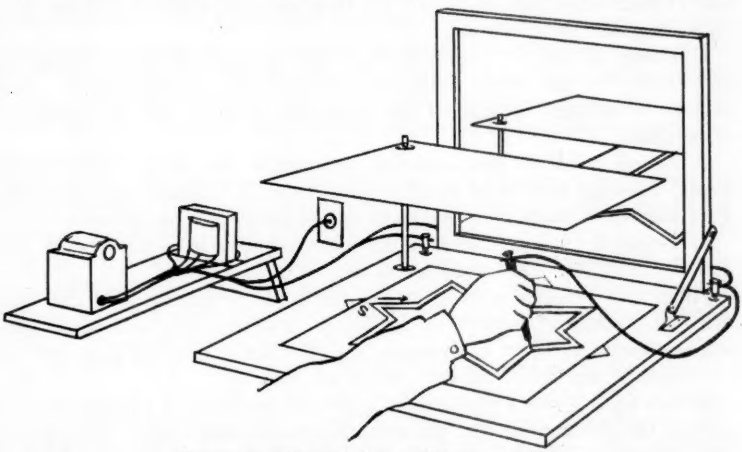


FIGURE Ia. View of the stabilimeter in use.

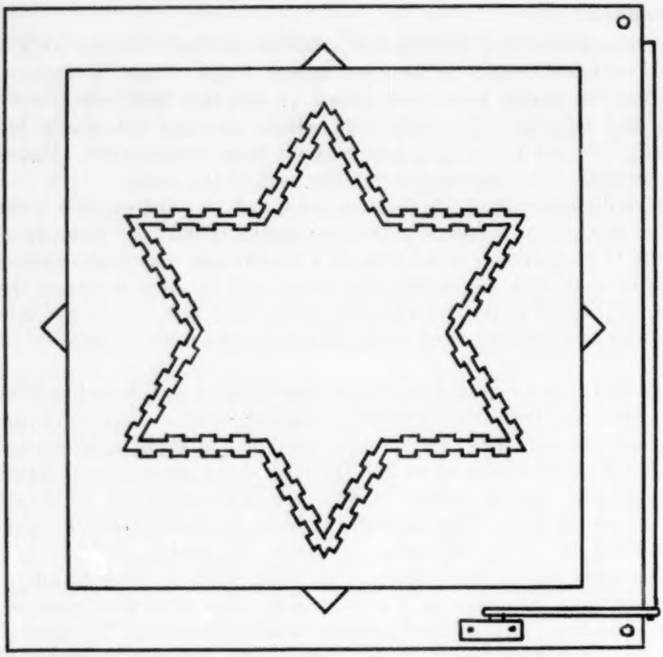


FIGURE Ib. Surface view of the stabilimeter.

tion of their scheduled practices. The time patterns explored and the designation for each group follow:

1. *Group X (MWF)* practiced three days a week, Monday, Wednesday, and Friday, for five weeks. They had a total of 15 practice periods. These days were selected because of the frequency of their use in educational scheduling.

2. *Group Y (M-F)* practiced five consecutive days of the week for five weeks. The total number of practice periods was 25. Monday through Friday were selected in order to approximate the typical education schedule.

3. *Group Z (SS)* practiced on the following days from the first: 1, 1, 2, 3, 5, 8, 13, 21, 34. This group had a total of nine practices.

Miller (4) reported a drop in superiority of performance of the group using his "additive" pattern on the fourth practice. This summation series was selected in an effort to eliminate this fall since it provides a greater concentration of practice in the early stages of learning.

In this study a day's practice for a subject consisted of tracing three circuits of the stabilimeter with ten seconds between each circuit. A single circuit consisted of tracing the irregular path of the six-sided star from the

starting point to the finish. Each contact of the metal stylus with the edge of the star shaped track was mechanically recorded as an error. The subject's daily score was the total of the time and errors for all three circuits.

The subjects were paced while using the instrument. This is a technique developed by Snoddy (6). In pacing, the time in seconds and the errors are kept approximately equal. This is accomplished by telling the subject to go faster when the number of seconds is greater than the number of errors, and by telling the subject to slow down when the number of errors exceeds the number of seconds that have been used. By this technique, it becomes possible to embed two dimensions, time and errors, into the measurement of the learning process. It should be noted that the pacing of the subjects is an extremely important feature of the experimental procedure. A considerable amount of skill and precision is necessary in properly utilizing the technique of pacing.

Lorge (3) found that performance was the best basis for equating groups for this particular laboratory skill. In this study the best single circuit score achieved by the individual on the first experimental day was the criterion used for assigning individuals to specific groups. As a result of the matching process, all three groups, each consisting of 32 subjects, had identical total scores and means.

A retention check was held for all subjects in each group two weeks following their final practice. The procedure for the retention check was exactly the same as the procedure on a regular practice day.

Presentation and Analysis of Data

Since group progress was the major concern of this study, the daily mean for the group was considered as the measure of group performance. Because improvement in performance results in the mean becoming lower, the reciprocal of the mean (1000/M) has been used on the figures in order to convert from descending to ascending curve. In the interpretation of the data, it was assumed that the degree of learning may be measured by differences in performance levels.

Both inter- and intra-group performances were compared by the use of standard statistical measures, including the critical ratio, and differences were considered significant at the .05 level.

Figure II shows the performance level of the three experimental groups throughout the total time of the experiment. In this graph the reciprocal of the mean score for each group has been plotted on the ordinate axis while calendar days are shown on the other axis.

Figure III shows a comparison of the performance level of the three experimental groups on the basis of the actual number of practice days. A comparison of the groups' performance at the time of the retention check is also indicated. The reciprocal of the mean for each day's performance has been plotted against the actual number of days that each group practiced.

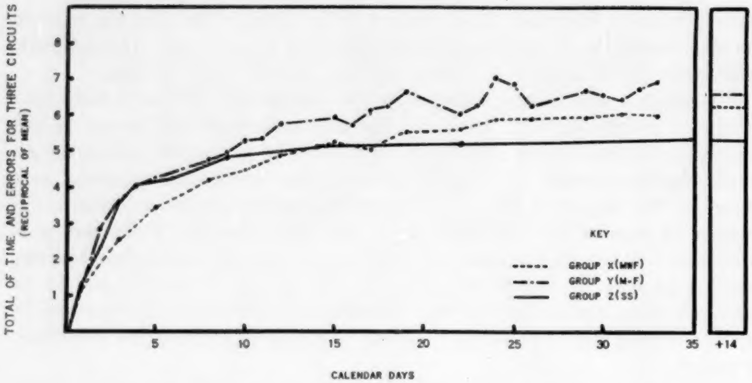


FIGURE II. Comparison of group performances and retention check on a calendar day basis.

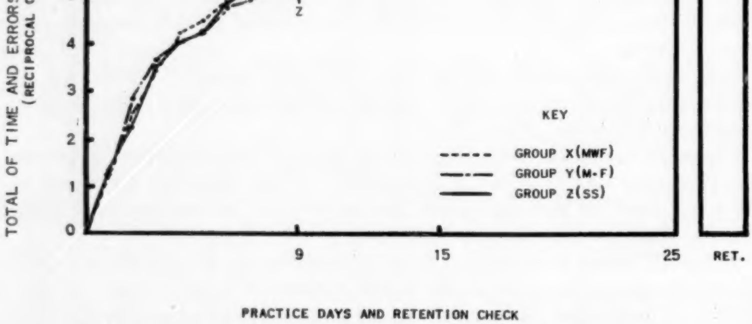


FIGURE III. Comparison of group performances and retention check on basis of experimental practice days.

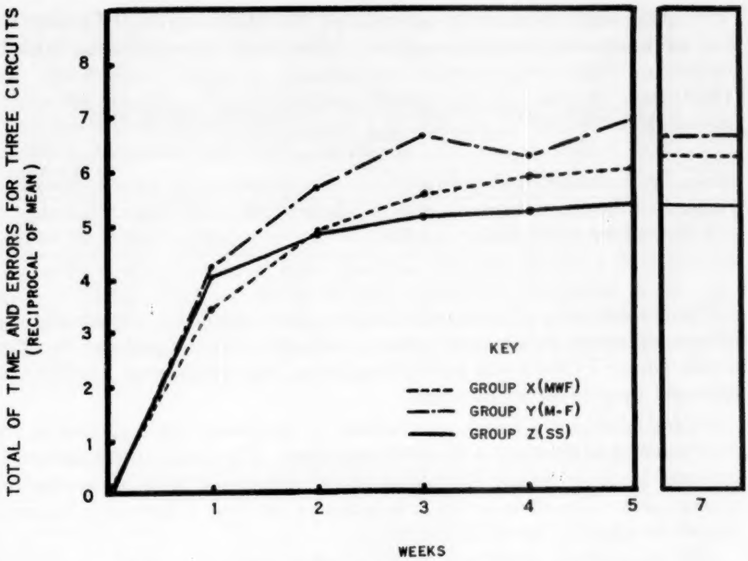


FIGURE IV. Comparison of group performances and retention check on basis of final performance of each week.

A comparison was made of the performance of the three experimental groups after each had had nine days of practice. The data appear in Table 1.

TABLE 1.—COMPARISON OF GROUPS X, Y, AND Z FOLLOWING NINE DAYS OF PRACTICE

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	179.69	32.74	5.79	8.47	8.63	0.98	5.57
Y(M-F)	188.16	36.24	6.41				5.31
Y(M-F)	188.16	36.24	6.41	2.78	9.11	0.30	5.31
Z(SS)	185.38	36.66	6.48				5.39
X(MWF)	179.69	32.74	5.79	5.69	8.69	0.65	5.57
Z(SS)	185.38	36.66	6.48				5.39

It should be noted that the ninth practice came at various times, depending upon the time pattern that the particular group was following. There were no statistically significant differences between any of the three groups.

It was possible to compare only groups X(MWF) and Y(M-F) after 15 days of practice. The data assembled for this comparison appear in Table 2.

TABLE 2.—COMPARISON OF GROUPS X AND Y FOLLOWING FIFTEEN DAYS OF PRACTICE

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	166.19	24.52	4.33	16.31	6.37	2.56	6.02
Y(M-F)	149.88	26.39	4.66				6.67

The critical ratio of 2.56 indicated that there existed a statistically significant difference between the means at this point of comparison. In other words, Group Y(M-F) was performing better than was Group X(MWF) as indicated by a lower mean score.

Figure IV shows a weekly comparison of the group performances at the practices and at the time of the retention check. The point of comparison is indicated by the reciprocal of the mean for each group on the final practice day of each week. The retention check is indicated as Week 7 because it occurred two weeks after the terminal practice.

In looking at the graph one should remember the range in the number of practices. Group Y(M-F) practiced five days a week for the entire five weeks. Group X(MWF) practiced three times a week. Group Z(SS) started out by practicing four times the first week and practiced progressively less as the weeks passed.

It is interesting to note that the lines for Group X (MWF) and Group Y(M-F) cross at the end of the second week. Each of the groups had had six practices at this point although their schedules had differed.

The data on the performance level of the three groups at the end of the first experimental week appear in Table 3. At this time Group X(MWF) had had three practices, Group Y(M-F) had completed five practices, and Group Z(SS) had had four practices.

TABLE 3.—COMPARISON OF GROUPS X, Y, AND Z FOLLOWING ONE WEEK OF THE EXPERIMENTAL PERIOD

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	289.63	61.92	10.95	54.63	8.21	6.65	3.45
Y(M-F)	235.00	40.92	7.23				4.25
Y(M-F)	235.00	40.92	7.23	8.47	12.56	0.67	4.25
Z(SS)	243.47	58.11	10.27				4.11
X(MWF)	289.63	61.92	10.95	46.16	15.01	3.07	3.45
Z(SS)	243.47	58.11	10.27				4.11

The level of performance of the three groups at the end of one experimental week appeared to be related to the number of practices that the group had had. The standard deviations also indicated decreasing variability in proportion to the number of performances. There was, however, no statistically significant difference between Groups Y(M-F) and Z(SS), although Group Y(MF) had practiced one additional time.

The differences in performance levels following four experimental weeks of practice are assembled in Table 4. At this time there existed a range of eight to 20 in the number of practices that each group had experienced.

In spite of a difference of eight practices between Group X(MWF) and Group Y(M-F), there was no statistically significant difference in the performance level of these two groups. Since the month is a common unit, the equality of performance that the two patterns produced should be a point of consideration in deciding between them.

TABLE 4.—COMPARISON OF GROUPS X, Y, AND Z FOLLOWING FOUR WEEKS OF THE EXPERIMENTAL PERIOD

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	169.47	31.79	5.62	10.22	8.14	1.25	5.90
Y(M-F)	159.25	33.33	5.89				6.28
Y(M-F)	159.25	33.33	5.89	31.16	8.19	3.80	6.28
Z(SS)	190.41	32.21	5.69				5.25
X(MWF)	169.47	31.79	5.62	20.94	8.00	2.62	5.90
Z(SS)	190.41	32.21	5.69				5.25

The comparisons which were made at the end of the fifth experimental week appear in Table 5. It should be noted that this was also the final practice for each of the groups.

TABLE 5.—COMPARISON OF GROUPS X, Y, AND Z AT TERMINAL PRACTICE WHICH OCCURRED AT END OF FIVE WEEKS OF THE EXPERIMENTAL PERIOD

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	166.19	24.52	4.33	21.72	6.32	3.44	6.02
Y(M-F)	144.47	26.03	4.60				6.92
Y(M-F)	144.47	26.03	4.60	40.91	7.95	5.15	6.92
Z(SS)	185.38	36.66	6.48				5.39
X(MWF)	166.19	24.52	4.33	19.19	7.80	2.46	6.02
Z(SS)	185.38	36.66	6.48				5.39

A comparison was made of the differences in performance levels of the three experimental groups on the day of the retention test. The data appear in Table 6.

TABLE 6.—COMPARISON OF RETENTION PERFORMANCE OF GROUPS X, Y, AND Z

Groups Compared	Means	S.D.	S.E.m	Dm	S.E.d	C.R.	Reciprocal 1000/M
X(MWF)	160.16	26.22	4.63	8.91	6.27	1.42	6.24
Y(M-F)	151.25	23.90	4.22				6.61
Y(M-F)	151.25	23.90	4.22	35.63	7.15	4.98	6.61
Z(SS)	186.88	32.67	5.77				5.35
X(MWF)	160.16	26.22	4.63	26.72	7.40	3.61	6.24
Z(SS)	186.88	32.67	5.77				5.35

The critical ratio of 1.42 for Group X(MWF) and Group Y (M-F) indicates that there was no significant difference between the performance on the retention test of these two groups. Group Y(M-F) had had ten more practice days than Group X(MWF). The similarity in performance is interesting to note because at the end of five weeks, or at the terminal practice, Group Y(M-F) was performing better than Group X(MWF).

The performance level of each group on the retention test was compared with the group's own performance on the final day of practice. These comparisons are shown in Table 7.

TABLE 7.—COMPARISON OF FINAL PRACTICE WITH RETENTION PERFORMANCE OF GROUPS X, Y, AND Z

Groups Compared	Time of Comparison	Means	S.D.	S.E.m	Dm	S.E.d	C.R.
X(MWF)	Prac. 15	166.19	24.52	4.33	—6.03	6.35	0.95
	Reten.	160.16	26.22	4.63			
Y(M-F)	Prac. 25	144.47	26.03	4.60	6.78	6.25	1.08
	Reten.	151.25	23.90	4.22			
Z(SS)	Prac. 9	185.38	36.66	6.48	1.50	8.68	0.17
	Reten.	186.88	32.67	5.77			

The critical ratios indicate that, for each group, there were no statistically significant differences.

Since Group Y(M-F) was the only group to practice beyond the fifteenth practice a comparison of their performance on the fifteenth and twenty-fifth day was made. This comparison appears in Table 8.

TABLE 8.—COMPARISON OF PERFORMANCES OF GROUP Y AT PRACTICES 15 AND 25

Groups Compared	Time of Comparison	Means	S.D.	$S.E._m$	D_m	$S.E._d$	C.R.
Y(M-F)	Prac. 15	149.88	26.39	4.66	−5.41	6.55	0.83
	Prac. 25	144.47	26.03	4.60			

The critical ratio of 0.83 indicates that the difference between the means is not statistically significant. A close examination of the data between these two points revealed that the performance of the group was unstable during this period. In Figure II it is apparent that during this period Group Y exhibited peaks of performance which were achieved by no other group.

Summary and Conclusions

The variety in the number of practices that each group had is important to remember in drawing conclusions from this study. In addition, since in each of the practice schedules followed there was a minimum of 24 hours between performances, they should be characterized as relatively distributed time patterns. Consequently, caution should be taken in comparing the results of this study with those obtained in investigations which involve massed practice. Furthermore, since the time interval between practices varied for each of the experimental groups, this factor must also be considered in drawing conclusions from the obtained results. However, the uniformity of the five experimental weeks of practice for the three groups presented a common practical situation in which the three selected time schedules of practice might be examined. Within this framework of reference, the following conclusions may be drawn from the data obtained in this study:

1. Since after nine practices there were no statistically significant differences in the performance levels among the three groups involved in this study, and since in this study the time periods between the first and ninth practices were 10, 18, and 34 days respectively, it seems reasonable to suggest that the total time span of the first nine practices (scheduled in this study) had a relatively insignificant effect on the degree of acquisition in the learning of the motor skill.

2. In a situation where 15 practices could be scheduled for learning motor skills, it would appear to be slightly more advantageous to hold group practice every day, Monday through Friday, for three weeks than to practice on Monday, Wednesday, and Friday for five weeks.

3. In learning this type of motor skill, the performance level at the end of one week tends to be approximately the same, whether the group practices on four or five consecutive days.

4. If either two or three weeks are available for the practice of a motor skill, it would appear that the performance at the end of these weeks would be best if there were a practice every day, Monday through Friday. Further-

more, the performance levels produced by the Monday, Wednesday, Friday schedule and the "Summation-Series" pattern would probably not differ significantly at the end of the second and third week, although at the end of three weeks the use of the "Summation-Series" pattern would have eliminated the necessity of two practices.

5. In learning a motor skill, there seems to be no significant difference in the level of achievement at the end of four weeks whether practice has been held on Monday, Wednesday, and Friday or Monday through Friday for the period.

6. If the performance of motor skills is to be judged at the end of five weeks, it appears that a practice schedule of Monday through Friday would have the tendency to produce significantly better results than either the Monday-Wednesday-Friday or the "Summation-Series" pattern.

7. If there are to be five weeks of practice in which to learn a motor skill and a span of two weeks before the performance will be judged, there will probably be no statistically significant difference in the performance, whether practice is held on Monday, Wednesday, and Friday or Monday through Friday. These two schedules of practice seem to produce better group performances than the use of the "Summation-Series" pattern.

8. The comparison of the performance level of each group at the final practice with its own retention level revealed no significant difference. This suggests the probability that the retention of the learning accrued from practice according to any of the three time patterns explored in this study was affected neither by the number of practices nor by their particular distribution.

9. Although the "Summation-Series" pattern appeared equally effective after nine practices, the over-all time utilized by the interpolated intervals must be considered in its evaluation. In a situation where the over-all time is not restricted, the "Summation-Series" pattern would be useful. The situation would also have to be one in which average performance level was acceptable, since the other two patterns explored in this study produced a higher level of skill at the end of the designated time of five weeks. The improvement during the last practices was slight. Two favorable characteristics of the performance of Group Z(SS) in this experiment were that there was an improvement apparent with each practice and that the retention level appeared comparatively stable. This suggests the possibility that these may be two characteristics of this particular "Summation-Series" pattern.

10. At the end of four weeks and at the time of the retention test there was no statistically significant difference between Groups X(MWF) and Y(M-F). Therefore, it seems reasonable to conclude that these two time patterns are at least equally effective. Since Group Y(M-F) had practiced ten additional times, one might also conclude that the Monday-Wednesday-Friday schedule was more efficient.

11. Group Y(MF) exhibited levels of performance that were attained by

no other group. Group Y's performance during the latter half of the experiment fluctuated considerably. The appearance of this instability tends to support Snoddy's (7) concept of secondary mental growth. It is also possible to conclude that the 24-hour interval, after three weeks, may have produced inhibition detrimental to performance. The tremendous difference in the amount of the time spent in practice would appear justifiable only when unpredictable peaks of performance are considered valuable.

In summary it may be stated that the findings of this study do not indicate the superiority of any one of the time patterns that were investigated. At the last point at which the performances of the three groups were compared on the basis of equal days of practice, there were no statistically significant differences among the performance levels of the three groups. At the end of 15 practice days, which provided the last comparison of Group X(MWF) and Y(M-F) based on an equal amount of practice, Group Y(M-F) was performing significantly better than Group X(MWF). However, at the time of the retention test, Group X(MWF) was performing as well as Group Y(M-F) even though Group Y(M-F) had had ten additional practices.

Many interesting and significant characteristics of the three patterns became obvious in examining the data. It would appear that a choice of one of the examined patterns would depend upon the particular objectives of the program in which the practices were being scheduled.

References

1. Dore, L. R., and Hilgard, E. R. "Spaced Practice and the Maturation Hypothesis." *Journal of Psychology* **4**: 245-59; Oct. 1937.
2. Lashley, K. S. "The Acquisition of Skill in Archery." *Papers from Department of Marine Biology*, Carnegie Institution of Washington **7**: 105-28; 1915.
3. Lorge, I. *Influence of Regularly Interpolated Time Intervals upon Subsequent Learning*. No. 438. New York: Teachers College, Columbia University, 1930.
4. Miller, A. G. *The Effect of Various Interpolated Time Patters on Motor Learning.* Unpublished Doctoral dissertation. Boston: Boston University, 1948.
5. Murphy, H. H. "Distribution of Practice Periods in Learning." *Journal of Educational Psychology* **7**: 150-62; July 1916.
6. Snoddy, G. S. "Learning and Stability." *Journal of Applied Psychology* **10**: 1-36; Mar. 1926.
7. ———. *Evidence for Two Opposed Processes in Mental Growth.* Lancaster, Pa.: Science Press, 1935.
8. Young, O. G. "The Rate of Learning in Relation to Spacing of Practice Periods in Archery and Badminton." *Research Quarterly* **25**: 231-43; May 1954.

(Submitted 4/4/58)

Effectiveness of Warm-Up Exercise in Junior High School Girls[1]

BETTY A. PACHECO
Bancroft Junior High School
San Leandro, California

Abstract

Five trials per day on the Jump and Reach test were given to 166 girls of grades 8 and 9 to determine if performance was improved by three minutes of vigorous running in place as a warm-up exercise. A balanced order test-retest experimental design was used, with the warm-up administered on Day 1 for half the subjects and on Day 2 for the others. A significant improvement of 4.83 percent was observed, caused by the warm-up. A practice effect caused the Day 2 scores to average 6.51 percent higher than the Day 1 scores. There was also a significant improvement from practice, from trial to trial within each day. The average test-retest reliability coefficient was $r = .868$. There were indications that reliability could be improved by averaging the scores obtained on several days but could not be improved very much by giving more trials in a single day test. Ninth-grade girls averaged 13.4 percent higher in their jumps than eighth-grade girls. The large practice effect needs consideration in the interpretation of norms.

THE POSSIBLE VALUE of using warm-up exercises for improving the physical performance of the younger age groups has received very little investigation. The single study concerned with young subjects (2) failed to detect any value from preperformance exercise. Using repetition of a 50-yd. sprint as warm-up for consecutive sprints that followed, there was no improvement in speed of running in Hipple's sample of 28 eighth-grade boys. These findings raise the question of whether adolescents or younger children may possibly have such resilient tissue structures that warm-up is unnecessary. It is also possible that the negative results might have been a reflection of the inadequacy of the specific type or amount of warm-up, or of the nature of the test performance, or of the rapid fatigue of younger subjects.

In the case of young adult males, the value of preperformance exercise as a warm-up procedure is now well established. The writer found statistically significant positive effects on the jumping performance of college men using several different types of preperformance exercises in two controlled experiments (6). Subsequently, Michael and Skubic (5) have reported significant improvement using a different type of warm-up exercise and performance, and Swegan, Yankosky, and Williams (8) have observed that repetitive movement used as the warm-up produces faster performance of that movement in the case of arm flexion. More recently, Thompson (9) has found that the "formal" type of warm-up exercise improves performance. The earlier work has been reviewed elsewhere (6).

---

[1]The writer is indebted to Franklin M. Henry of the University of California for encouragement and for help with the statistical analysis.

The effectiveness of preliminary exercise in improving the performance of women has been given little attention. There has been only one study of female subjects, insofar as the writer is aware. Skubic and Hodgkins (7) tested 31 college women physical education majors using three types of warm-up procedures and found no significant effect. However, the preliminary exercises that were used were quite mild and of short duration, which may possibly explain the absence of positive results.

In view of the current status of the warm-up problem as outlined above, it has appeared desirable that the present investigation should be concerned with studying the performance of young girls with respect to the possible value of warm-up exercises. In view of the positive results obtained with college men (6), it has seemed best to use vertical jumping as the test performance. However, the automatic recording device used in that study was not readily available. It was decided, therefore, to use the jump and reach test. There was a specific interest in finding out if this test was sufficiently accurate to measure differences of only a few percent, as might result from application of an experimental modality such as warm-up. Although the jump and reach is widely used in test batteries, the information necessary to evaluate it for the present purpose, particularly considering the type of subjects, could not be found. The experiment is, therefore, exploratory in nature.

Since the jump and reach test is coming into widespread use in the junior high school programs as a result of current interest in physical fitness, it has seemed important to investigate the possible presence of significant practice effects during its administration. It cannot be determined from the literature if such an effect exists, or if it is large enough to cause concern. It is known that there is a practice effect in the case of men subjects performing the Sargent Jump test and that it can be reduced to a small amount by modifying the original method in the direction of reducing the arm movements to a minimum (4, 6). Young girls might exhibit a much greater practice effect, and the necessity of using the arms in the jump and reach might also contribute, since it would be necessary to learn how to coordinate the arm and leg actions in order to make the most effective jump.

Method

Experimental Design. One hundred sixty-six girls of grades 8 and 9 were divided into two groups, A and B, which differed in that the testing in B was done in reverse order to balance out learning, conditioning, or other unforeseen factors. The division was random, with some adjustment so that half the members from each grade were in each group.

The subjects were tested on two separate days, at the same hour both times, but with an interval of one week. On the first day, Group A did five vertical jumps with a half-minute rest between each jump. After jumping, the subjects did three minutes of running in place. On the second test day

this group did this three-minute exercise as a warm-up procedure, prior to the vertical jumps. Approximately one and one-half minutes elapsed between the conclusion of the running and the first vertical jump.

Group B was tested in the reverse order. On the first day of testing, these subjects ran three minutes in place as a warm-up, and then did five vertical jumps with a half-minute rest between each jump. On the second day of testing, they jumped five times and then did the three-minute exercise.

The students were encouraged to jump as high as possible each time. No practice jumps were allowed. The instructions throughout the testing stressed that they should not change their jumping technique. No special apparatus was used to measure the height of the jump. The true purpose of the preliminary exercise was not revealed. In order to help secure psychological control, pulse counts were taken after the running in most cases. This aided in disguising the true purpose of the testing.

Procedure. The subject held a piece of chalk one and one-half inches long between the thumb and index finger. Facing the gymnasium wall, she reached overhead and made a short horizontal chalk mark (see Figure I). Her position was checked to ensure that the toes were touching against the wall and that the arm was extended straight above the head. She then turned her preferred side to the wall and jumped.

Warm-Up Exercise. Running was selected as the warm-up activity, because in a previous study by the writer it was shown to have a greater effect upon jumping performance than two other types of warm-up exercises. Also, running is frequently the method of warm-up used by students before engaging in competitive performance and is often used as a preliminary exercise in physical education classes. During the three-minute run, the girls were urged to use their arms as well as their legs. No special pace was required for the running. Each subject chose a comportable pace which she felt could be maintained for the three-minute period and which resulted in a good workout. If the pace slowed down noticeably or the arms were not moved with adequate vigor, the girl was instructed to work harder at the task.

Results

Matching of Subgroups. McNemar (3) has shown that exact matching of subgroups is not necessary to secure a valid interpretation of an experimental effect if the subgroup differences before treatment are within the limits of random sampling. Each subgroup of individuals in this study had both the experimental and control conditions (only the order of treatment differed). In the eighth-grade subjects, Group A jumped on the average 1.151 cm. higher than B ($t=1.29$); in the ninth-grade subjects, Group A jumped 1.275 cm. less than B ($t=0.63$). Both of these differences are well within the limits of random sampling. The average scores for all subjects are almost identical for Groups A and B; the difference is only 0.07 cm.

Warm-Up Effect—Daily Averages. Using the differences between the five-

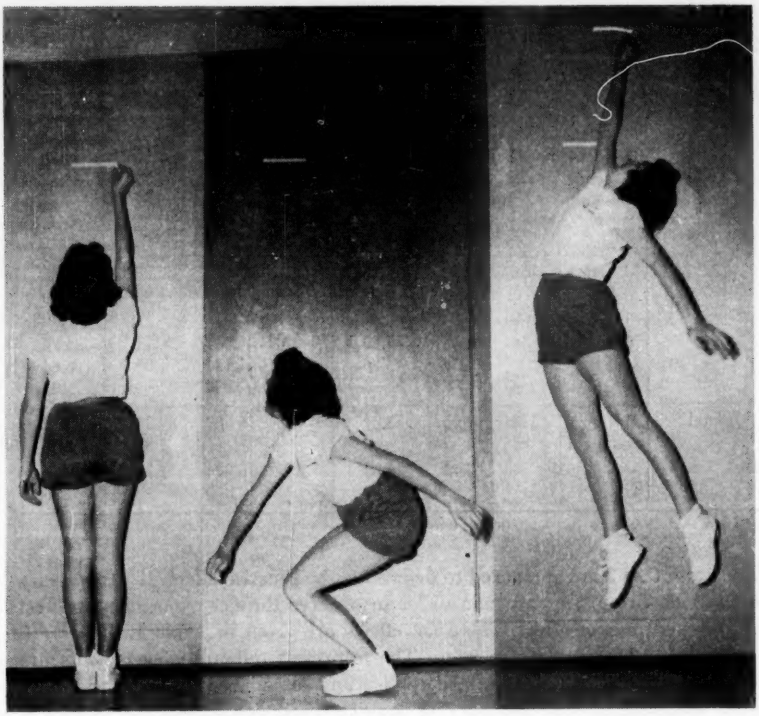


FIGURE I. Performance of the Jump and Reach test.

trial daily averages under experimental and control conditions, the average gain due to warm-up is 1.56 cm. (t=4.52) for the entire group of 166 girls. For the eighth-grade subjects alone, the gain is 1.42 cm. (t=3.48); for the ninth-grade subjects it is 1.82 cm. (t=2.87). All these differences are significant at the 1 percent level.

Practice Effect—Daily Averages. For the 166 subjects, the average practice effect obtained by subtracting the mean scores on Day 1 from those of Day 2 is 2.02 cm. (t=6.13). In the eighth-grade subjects it is 1.75 cm. (t=4.42). In the ninth-grade subjects the practice effect is 2.61 cm. (t=4.36). Here again the differences are significant at the 1 percent level.

A detailed breakdown of the daily average scores is given in Table 1. It should be noted that the experimental condition scores for Group A include also the practice effect, whereas its influence is added to the control scores in Group B. A comparable situation occurs with respect to the differential influence of the experimental effect on Groups A and B when the scores of

TABLE 1.—MEAN JUMP SCORES AND RELIABILITY COEFFICIENTS FOR FIVE-TRIAL AVERAGES

Group	Statistic	Eighth Grade			Ninth Grade			Average[a]
Practice effect		N	Day 1	Day 2	N	Day 1	Day 2	
"A"	M (cm.)	56	31.29	34.45	27	33.61	38.05	34.35
"B"	M (cm.)	56	31.29	31.62	27	36.71	37.50	34.28
Total	M (cm.)	112	31.29	33.04	54	35.16	37.77	34.32
	Net gain			1.75			2.61	2.18
Exp. effect		N	Control	Exp.	N	Control	Exp.	
"A"	M (cm.)	56	31.29	34.45	27	33.61	38.05	34.35
	σ		5.59	6.80		5.83	7.92	6.60
	r		.928			.851		.898
"B"	M (cm.)	56	31.62	31.29	27	37.50	36.71	34.28
	σ		5.35	5.26		7.84	7.65	6.64
	r		.721			.898		.831
Total	M (cm.)	112	31.45	32.87	54	35.56	37.38	34.32
	Net gain		1.42				1.82	1.62
	σ		5.47	6.08		6.91	7.79	6.62
	r		.858			.879		.868

[a] Note that these are not weighted averages.

Day 1 and Day 2 are compared to determine the practice effect. It is necessary to combine Groups A and B in order to observe the net experimental effect and the net practice effect. These net effects are given in Table 1.

Reliability of Five-Trial Scores. The test-retest reliability coefficients are shown in Table 1. Because of the complications introduced by the A, B—B, A order of testing, it was necessary to compute these coefficients for Groups A and B separately and then average them by the Z-transformation. The influence of the experimental effect would reduce the reliability somewhat. This will be discussed later, in the section on single trial comparisons. It is doubtful, however, if there are any significant differences among these correlations. While it is true that the difference between the reliability coefficients for Groups A and B in grade 8 ($r=.928$ and $.721$) is statistically significant, since t_z is 2.54, one is not justified in picking out extreme differences in this manner. The average correlations for the two grades combined do not differ significantly between Groups A and B ($r=.898$ and $.831$: $t_z=1.71$), and the average correlations for the two grades, averaging A and B, are even closer together ($r=.858$ and $.879$).

Single Trial Results

Practice Effect. Figure II shows that there is a very large practice effect among the five trials within each day. This effect is statistically significant within each school grade and within each day and condition. This is established by the eight variance analyses summarized in Table 2. The F-ratios for

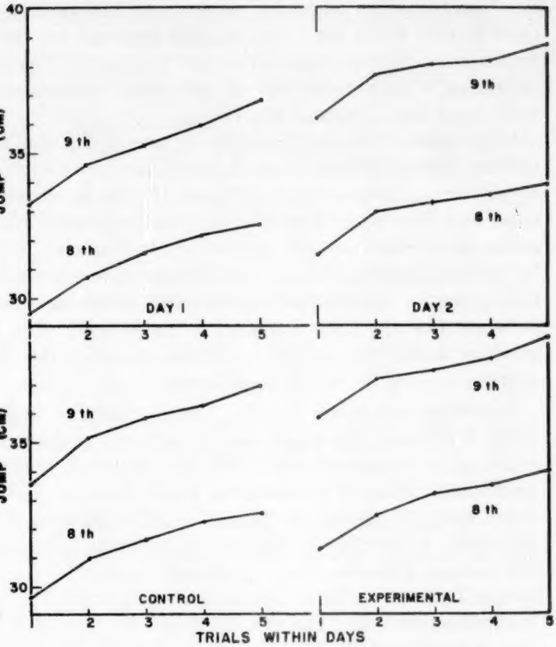


FIGURE II. Trial-by-trial practice effect within days and within treatments.

TABLE 2.—VARIANCE ANALYSES OF TRIALS WITHIN DAYS AND TREATMENTS

Condition	Statistic	Source of Variance							
		Eighth Grade				Ninth Grade			
		Total	Subjects	Trials	Error (E)	Total	Subjects	Trials	Error (E)
Day 1	df	279	55	4	220	134	26	4	104
Control	MS	35.275	158.923	81.490	3.522	38.865	176.991	55.618	3.689
F-ratio	MS/E		45.12[a]	23.14[a]			47.98[a]	15.08[a]	
Exper.	MS	33.107	141.013	93.013	4.573	62.999	304.250	43.688	3.416
F-ratio	MS/E		30.84[a]	20.54[a]			89.05[a]	12.79[a]	
Day 2	df	279	55	4	220	134	26	4	104
Control	MS	32.699	145.474	64.695	3.923	76.780	375.555	34.790	3.702
F-ratio	MS/E		37.08[a]	16.49[a]			101.45[a]	9.40[a]	
Exper.	MS	51.367	235.136	38.120	5.665	65.541	326.146	18.750	2.189
F-ratio	MS/E		41.51[a]	4.96[a]			149.97[a]	8.56[a]	

[a] All F-ratios are statistically significant at the 1 percent level.

the "practice effect by trials" range from 4.96 to 23.14, and the smallest of these is well above the value of 3.40 required for the one percent level of significance. It is also noted that the F-ratios for "subjects" are quite large, indicating a high reliability of individual differences for successive trials withn each day's series of five jumps.

Improvement between trials due to practice is just as evident on days involving preperformance warm-up exercises as on control days. These curves are shown in the lower half of Figure II. This is an important result, because there had been some fear that the relatively large amount of warm-up exercise might cause enough fatigue to interfere with performance, particularly in the eighth-grade girls. If such fatigue were present, it would be expected to depress the trial-by-trial improvement under experimental conditions. The facts are that the improvement for the younger girls is almost exactly the same as under the control conditions, and for the ninth-grade girls it is nearly the same as control conditions.

Reliability of Single Trials. These reliability coefficients are given in Table 3. The average coefficient for adjacent trials is $r=.897$ for the eighth grade girls, compared with .953 for the ninth-grade girls ($t_z=1.82$, not significant). A lower reliability is found between the first and tenth jumps, which were, of course, performed on different days. The coefficient for the older girls is $r=.823$ in this case, a significant drop from .953 ($t_z=3.53$). The younger girls also drop significantly from .897 for adjacent trials to .673 for the Trial 10 vs. Trial 1 condition ($t_z=5.22$). The latter coefficient, .673, is considerably lower than the .823 found for the older girls ($t_z=2.43$, significant at the 2% level).

The size of the standard deviation of the differences given in Table 3 is of interest. For the younger girls it is 5.44 between Trials 1 and 10, nearly as large as the figure 5.78 for their distribution of scores on Trial 1. This indicates that these girls change as much within themselves over the period of ten trials as they differ between each other on the first trial. Such a finding is, of course, implied by the relatively low reliability coefficient relating Trials 1 and 10 ($r=.637$).

Not all of the differences between Trials 1 and 10 can be charged to the practice effect, although it is the major factor. In the eighth-grade girls, the practice effect is responsible for 4.46 cm. of improvement, and the influence of the warm-up exercise is 1.46 cm. The latter augments the observed improvement in scores in Group A and decreases it in Group B (because it preceded Trial 1 in the latter case). The corresponding figures for the older girls are 5.44 cm. for the practice effect and 1.88 cm. for the warm-up exercise. In other words, the experimental treatment (warm-up) has an influence that is only 33 percent as large as the practice effect between Trials 1 and 10. In the absence of more direct information, it can be inferred that the relative effects of practice and warm-up on the variability of the differences is proportional to the effects on the differences.

TABLE 3.—MEAN JUMP SCORES AND RELIABILITY COEFFICIENTS FOR SINGLE TRIALS

Group	Statistic	Eighth Grade				Ninth Grade			
	Trial	1	2	9	10	1	2	9	10
"A"	M (cm.)	29.47	30.83	34.68	35.39	31.44	33.10	38.20	38.86
	σ	5.88	5.82	7.26	6.95	5.56	6.00	8.21	7.71
	Diff. between trials	(5.92)[a]	1.37		0.71	(7.42)	1.66		0.66
	σ difference	(5.56)[a]	2.25		3.25	(3.20)	2.18		1.91
	r	(.706)[a]	.926		.897	(.774)	.932		.974
"B"	M (cm.)	29.43	30.58	32.46	32.43	35.04	36.10	38.17	38.70
	σ	5.68	5.61	5.92	5.56	8.35	7.78	8.89	8.56
	Diff. between trials	(3.00)	1.15		−0.03	(3.66)	1.06		0.53
	σ difference	(5.30)	2.99		2.65	(4.45)	2.34		3.20
	r	(.556)	.859		.896	(.862)	.960		.933
Total	M (cm.)	29.45	30.71	33.57	33.91	33.24	34.60	38.19	38.78
	σ	5.78	5.72	6.62	6.30	7.10	6.94	8.56	8.14
	σ difference	(5.44)	2.65		2.96	(3.88)	2.26		2.63
	r	(.637)	.898		.896	(.823)	.948		.957

[a] Entries in parentheses refer to differences and correlations between Trials 1 and 10. Other differences and correlations are between adjacent trials (1-2 or 9-10). Correlations in this table (as well as Table 1) have been averaged by the z-transformation. Standard deviations have been averaged as σ^2.

Comprehensive Variance Analysis. In order to secure an over-all view of the net role of the various factors, a comprehensive variance analysis of the combined data has been made. It is summarized in Table 4. It may be noted that since the F-ratios for "practice effect within days" and for "experimental effect" (warm-up) have one degree of freedom in the numerator, the square root of F is the equivalent of the corresponding t-ratio. For the practice by days, this t-ratio is 4.78 for the eighth-grade girls and 4.12 for the ninth-grade

TABLE 4.—VARIANCE ANALYSIS OF ALL FACTORS

Source of Variance	Eighth Grade			Ninth Grade		
	df	MS	F-ratio	df	MS	F-ratio
Total Sum of Squares	1119	39.778	——	539	63.652	——
Days of practice	1	854.180	22.82	1	919.380	16.91
Exp. effect	1	562.170	15.02	1	449.350	8.27
Interaction[a]	110	41.998	9.50	52	54.359	16.73
Individual differences	110	301.269	68.15	52	551.998	169.90
Trials within days	8	135.883	30.74	8	74.334	22.88
Standard error of meas.	880	4.421	——	416	3.249	——

[a] This interaction, "subjects by days within orders," serves as the error term for days of practice, and also for the experimental effect. For individual differences (i.e., subjects within orders) and for the practice effect of trials within days, the error term is designated Standard Error of Measurement. It is the interaction term "subjects within orders by trials within days." The other interactions have not been computed, since they are not used in the analysis. All F-ratios are statistically significant at the one percent level.

group. For the warm-up effect the figures are 3.88 and 2.88. These values are slightly higher than the t-ratios calculated earlier by the conventional method, because the variance analysis yields more efficient statistics. The "trials within days" practice effect is highly significant, as was the case in Table 2. The high reliability of the individual differences between adjacent trials shown in other tables is confirmed by the very high F-ratio of 169.9 for individual differences.

Age Differences. The sample of ninth-grade girls in this study jumped 13.4 percent higher than the eighth-graders; this is a significant difference ($t=3.80$). The data of Table 4 reveal that the variability of the "true scores" of individuals is 83 percent greater for the older girls ($F=1.83$, significant at the 1% level); the error variance, however, is 36 percent smaller ($F=1.36$, also significant at the 1% level).

Discussion

The warm-up effect caused by stationary running was highly significant in both age groups, i.e., both experiments gave positive results. Failure of others to find this effect may possibly be explained by failure to use a sufficient amount of exercise. The results of Asmussen and Boje (1) show that a very large amount of preliminary exercise was required in order to produce appreciably large improvement in performance. In the present study the subjects complained of being fatigued as a result of the preliminary exercise, illustrating that the warm-up exercise was relatively severe.

These junior high school girls showed a score increase of 4.8 percent due to the warm-up. This is considerably less than the 7.8 percent improvement found in college men in the earlier study (6). Several possible reasons may be offered in explanation. It may be that the difference in the amount of anatomical development and maturation is responsible. Adolescents are considered to be much more flexible than adults. Also, it may be that females do not derive as much benefit from warm-up as males do. There is no controlled experimental evidence on this point.

The large amount of the trial-by-trial practice effect is of particular importance because it implies that norms for the jump and reach test for girls of the age-range studied cannot be used indiscriminately. The scores obtained will be high or low, depending not only upon ability, but also to a large extent upon how much experience and practice the subjects have had. The question of how much practice is required for learning to reach a stable plateau is one that needs further investigation.

Finding equal amounts of trial-by-trial practice effect for both control and experimental conditions (Figure II) implies that it is unlikely that this would occur if the upward trend of the curve were caused by the warm-up effect of one jump upon the one that follows it. The average practice effect within days was 2.8 cm. for the eighth-graders and 3.0 cm. for the ninth-graders. The average warm-up effect between control and experimental conditions was

1.4 and 1.8 cm. for the younger and older girls respectively; it was caused by a considerable amount of warm-up exercise. It does not seem reasonable that much smaller amount of exercise, as represented by only four jumps, could cause twice as much additional imrovement; or for that matter, that the improvement from the possible warm-up effect of the jumps would be so nearly the same, whether or not it was preceded by the prejump running in place.

Another point that bears on this problem is that college men, while showing some improvement within a series of five serial jumps on the first day of practice, do not improve during the second or third day, although the opportunity for the self-warm-up would still be present (4).

This finding can be confirmed by further study of the writer's earlier investigation of warm-up by pre-exercise in 50 college men tested with a modified Sargent Jump (6). A variance analysis of those data, not previously reported, making use of the same method as was used to construct Table 2, yields an F-ratio of 5.01 for the "trials effect" in the control group and 3.17 in the experimental group on the Day 1 performances. Both are highly significant, since a ratio of only 2.46 meets the requirement for the 5 percent level of confidence. When the two groups are averaged, the heights of the five consecutive jumps on Day 1 are 47.6, 47.9, 49.1, 49.4, and 49.6 cm. While the practice effect is less than for the girls it is nevertheless significant. On Day 2, however, there is no significant improvement in successive trials, since the corresponding F-ratios are only 2.40 and 1.87. Also, as in the Merlino study (4), the average performance on Day 2 is not significantly better than on Day 1 ($t=1.19$). Since there is a real practice effect from trial to trial during the first day but only random fluctuations from trial to trial during the second day, even though the opportunity for "self-warm-up" is equally present on Day 2, it seems that the trials effect is a true practice effect and not a reflection of warm-up caused by the jumping performance itself.

The considerably larger trial-by-trial practice effect in the girls, and its continuation almost unabated on Day 2 (Figure II), may have two causes. The girls are young and relatively inexperienced. They improved 4.3 cm. in jumping ability, just between the eighth and ninth grades. They are probably more sensitive to practice than the college men. There is also a good possibility that the arm action that is a necessary part of the Jump and Reach test, but which is excluded from the modified Sargent Jump, involves a considerable element of learnable skill.

The standard error of measurement within days may be computed as the square root of the error variance given in Table 4. This error includes both the "real" measurement or observational error and the variation in performance of the subjects from trial to trial within each day's series of five jumps, which would be designated as the intra-individual variation within days. The size of the standard error of measurement is surprisingly small, namely 2.10 cm. for the younger subjects and 1.80 cm. for the ninth-

graders. Neither of these figures is significantly different from the standard error of measurement of 1.94 cm. found in college men tested with the Sargent Jump using the Henry apparatus (6).

The day-to-day intra-individual variation in performance is less in the men than in the girls. Using the average of five trials as the individual score, the results given in Table 1 show that when the subgroups were averaged the eighth-graders had a test-retest reliability of $r=.858$ and a standard deviation of 5.47 cm. The standard error of measurement for predicting one day's performance from another, computed by the usual method, is 2.06 cm. For the ninth-graders the corresponding figures were .879, 6.91, and 2.41. In the case of the college men, $r=.975$, S.D.=7.82, and the measurement error is computed to be 1.237 (6). The day-to-day variation is, therefore, significantly less (1% level) in the men than in the eighth-graders ($F=2.77$) or the ninth-graders ($F=3.79$) and does not differ significantly between the eighth and ninth graders ($F=1.37$). Since the measurement error for trials within days is as large for the men as for the girls, it seems likely that the greater day-by-day variation of the girls is not caused by differences in apparatus or method but is rather an intrinsic variability characterizing their age or sex. Their reliability coefficient, which is .868 for the average of the two grades, is not so high as would be desirable for measuring individual differences. In view of the small variation within days, it would not help much to give more trials per day. If two days were averaged, the Spearman-Brown correction predicts a reliability of .929, which might be considered satisfactory.

Summary and Conclusions

The vertical jumping ability of 166 girls from grades 8 and 9 of a junior high school was determined by the jump and reach test. The subjects made five jumps per day on each of two days one week apart. Using a balanced experimental design, the jumps on one of these days were preceded by three minutes of running in place as a warm-up exercise.

The average jump of the eighth-grade girls was 31.5 cm. without, and 32.9 cm. with, warm-up. The corresponding scores of the ninth-grade girls were 35.6 without and 37.4 with warm-up. The gain was statistically significant in both cases. The girls in the older group jumped significantly higher, and the variation in ability between individuals was significantly larger.

A large and significant practice effect was found between successive trials within each day and from the first day to the second, for both age groups. These results point to the need for further research and to the necessity for considering the practice effect in the establishment and use of norms.

The five-jump individual daily scores did not differ significantly in test-retest reliability as between the two age groups. Averaging the two groups, the coefficient of reliability was $r=.868$. The within-day error of measurement was no larger than was previously found for the vertical jumping performance of college men, but the between-day variability was 82 percent larger. The

relatively small within-day variability indicated that improvement in the reliability of the test would have to be secured by averaging the results of several days of testing rather than increasing the number of trials per day.

The following conclusions seem justified by the data:

1. Preliminary warm-up exercise such as running in place improves the vertical jumping performance of girls of junior high school age.

2. The "self-warm-up" effect of one trial upon the next consecutive trials is a negligible factor in this improvement.

3. The jump and reach test as used in this study has a relatively high reliability for successive trials within one day's series of jumps, but it is only moderately reliable in predicting one day's performance from another.

4. The vertical jumping performance of girls of junior high school age is markedly influenced by practice.

5. Jumping performance at this age improves considerably over the small age change represented by a single school grade.

References

1. Asmussen, E., and Boje. O. "Body Temperature and Capacity for Work." *Acta Physiologica Scandinavica* **10**: 1-23; 1945.
2. Hipple, J. "Warm-Up and Fatigue in Junior High School Sprints." *Research Quarterly* **26**: 246-47; May 1955.
3. McNemar, Q. "Sampling in Psychological Research." *Psychological Bulletin* **37**: 331-65; 1940.
4. Merlino, Lawrence. "Influence of Massage on Jumping Performance." *Research Quarterly* **30**: 66-74; Mar. 1959.
5. Michael, E.; Skubic, V.; and Rochelle, R. "Effect of Warm-Up on Softball Throw for Distance." *Research Quarterly* **28**: 357-63; Dec. 1957.
6. Pacheco, B. "Improvement in Jumping Performance due to Preliminary Exercise." *Research Quarterly* **28**: 55-63; Mar. 1957.
7. Skubic, V., and Hodgkins, J. "Effect of Warm-Up Activities on Speed, Strength, and Accuracy." *Research Quarterly* **28**: 147-52; May 1957.
8. Swegan, D.; Yankosky, G.; and Williams, J. "Effect of Repetition upon Speed of Preferred-Arm Extension." *Research Quarterly* **29**: 74-82; Mar. 1958.
9. Thompson, H. "Effect of Warm-Up upon Physical Performance in Selected Activities." *Research Quarterly* **29**: 231-46; May 1958.

(Submitted 6/29/58)

Ability of Deaf Swimmers To Orient Themselves When Submerged in Water[1]

DONALD A. PADDEN
Gallaudet College
Washington, D. C.

Abstract

A study was conducted of the effect of submersion upon the sense of orientation of deaf swimmers. The deaf swimmers were classified into groups of those who were born deaf, those who were meningitic, those deaf from other causes, and those deaf from unknown causes. They performed two orientation tests along with a group of normal hearing swimmers. One test was the submerging of subjects under water from a rotating board with the eyes open and the other test was similar but with the eyes blindfolded. A statistical analysis of the data indicated that the meningitis group made a statistically significant slower time in coming up to the surface than the other groups.

IN ORDER TO MAINTAIN equilibrium of the body, constant nervous activity is necessary. This must be regulated by the central nervous system, which in turn is stimulated by the messages received from the various sensory receptors that indicate the movements affecting equilibrium. The sensory receptors which make up the complex mechanism of equilibrium are the eyes, the organs of touch in the skin and the proprioceptors in the muscles and joints, and the vestibular apparatus of the ears. These sensory functions are all delicately coordinated in the maintenance of equilibrium.

Upon impairment of one of the three components, equilibrium is affected until the other components compensate for the deficiency and reestablish equilibrium. The inner parts of the ears contain the organs of hearing and equilibrium and because of the proximity of the auditory and vestibular structures, individuals suffering deficiencies in hearing sometimes show vestibular defects. For that reason, the deaf, as a group, are known to be deficient in the sense of equilibrium. This problem has intrigued scientists for many years, and they have studied deaf subjects with various tests such as tracing the body sway while standing still, recording electrically the length of the arc of the swing while standing on a balancing platform, observing the presence or absence of nystagmus or dizziness after being turned around on a rotating chair, and measuring distance while walking on rails.

In articles on the balancing ability of the deaf, James (5), Obersteiner (8), Morsh (7), Edward (4), Davis (3), and Best and Taylor (1) have pointed

[1] This study was done in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Physical Education, Recreation, and Health, University of Maryland. The writer is indebted to Benjamin H. Massey for assistance and guidance throughout the study.

out that the deaf are known to lose their sense of orientation if, while swimming, they become submerged in water. The statements apparently were based solely upon experiments conducted on lower animals such as the studies by Prentiss (9), Streeter (10), and Camis (2), or upon information given by the deaf themselves. Insofar as this investigator could determine, no investigation has been made as to how much difficulty deaf swimmers actually do encounter in finding their bearings when submerged in water. A study along these lines is important since such information is fundamental for all physical education dealing with the deaf.

Purpose

The purpose of this study was to determine by means of an improvised test the relationship of deafness to the sense of orientation when a deaf subject was submerged in water. The specific purposes were (a) to determine the relationship of deafness from various causes to the ability of deaf subjects to orient themselves under water as indicated by the time measured in seconds that it took each deaf subject submerged in water to locate the surface, (b) to determine if the sense of orientation so measured would be affected by having the deaf subjects blindfolded while being submerged, and (c) to determine whether the sense of orientation is influenced by a deficiency in equilibrium, both with eyes open and blindfolded.

Procedure

Subjects. The subjects selected to participate in the experiment were 105 deaf and hard of hearing male students enrolled at Gallaudet College, Washington, D. C., and 23 normal male students from neighboring schools and colleges. The subjects ranged in age from 15 to 30 years, in height from 63 to 78 inches, and in weight from 123 to 205 pounds. All subjects were in good health, and with the exception of deafness none of them suffered any gross physical defects. At the time of the experiment the deaf subjects were in the regular, required physical education program. To qualify as a subject each individual had to satisfy a minimal requirement of being able to swim.

Apparatus. The apparatus used in the experiment consisted of a 4-inch square wooden beam, 18 feet long, and a rotating board made of plywood in octagonal shape, measuring 42 inches in diameter and ¾ inch thick. At the middle of the beam a round steel pin was driven into the wood, leaving approximately 2 inches exposed upon which the board could be rotated. The testing took place in a swimming pool 35 feet long and 14 feet wide. The depth ranged from 4 feet at the shallow end to 8 feet at the deep end. The beam was placed across the width of the pool and over water which was 7½ feet deep. Then on the steel pin in the beam was laid the rotating board, the top of which was 2 feet above the water, corresponding to the height of a diving board.

Administration of Tests. The procedure consisted of two tests, one in which

the subjects were tested with their eyes open and the other with their eyes blindfolded. A woman's swimming cap pulled over the eyes served as the blindfold.

Initially the subject's age, height, and weight were recorded. He was then instructed to sit on the rotating platform in a squatting position, to wrap his arms firmly about his legs, and to bow his head (see Figures I and II). In order to disorient the subject, the platform was turned around once, twice, or three times in a clockwise, counterclockwise, or both directions. It was then tilted, throwing the subject into the water forward, backward, or to either his right or left side. The subject was instructed to keep the squatting position with the arms around the legs until fully immersed, after which he was to immediately start swimming for the surface (see Figures III and IV).

The eyes open and eyes blindfolded tests consisted of 20 trials each. The eyes open test was always administered before the eyes blindfolded test; in other words, all 20 trials of the eyes open test were given before all 20 trials of the eyes blindfolded test. This procedure was a matter of arbitrary choice for ease of administration since no advantage could be seen to be derived from administering the eyes blindfolded test first or alternating the eyes open and eyes blindfolded tests. In each test the submerging was repeated five times from each of the four positions, forward, backward, right side, and left side. The direction of tipping was varied so that the subject would have no prior knowledge as to which way he would be thrown. No two sequences of submerging subjects were the same as the investigator consciously tried to mix up the sequence and not to follow a set pattern.

Trials in which the subject let go the hold on his legs before hitting the water, touched the bottom of the pool, pulled off the blindfold while still under water, or did not attempt to swim under water were cancelled. The reasons for this were that (1) the subject could prevent himself from becoming completely submerged by allowing a greater area of body surface to spread over the water, (2) touching the bottom would become a fixed point of reference for orientation under water, (3) the subject would depend on vision for finding the surface, and (4) the body would rise due to the force of buoyancy if he remained quiet, thus providing a cue as to the direction of the surface.

The subject was timed from the moment any part of his body hit the water until his head appeared above water. Three stop watches were used in timing the trials, and the average time of the three for each trial was used as the score. The investigator administered the submerging phase of the testing program while the deaf and hard of hearing subjects themselves took turns at keeping time and recording scores for fellow subjects. The normal hearing subjects, however, were not permitted to keep time as it was believed that the response of a hearing timer to cues might be different from that of a nonhearing timer. One of the principles of perceptual speed and quick responses

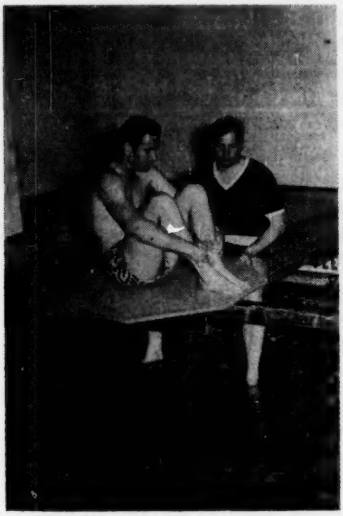


FIGURE I. A subject with eyes open waiting to be thrown into the water.

FIGURE II. A subject with eyes blindfolded waiting to be thrown in.

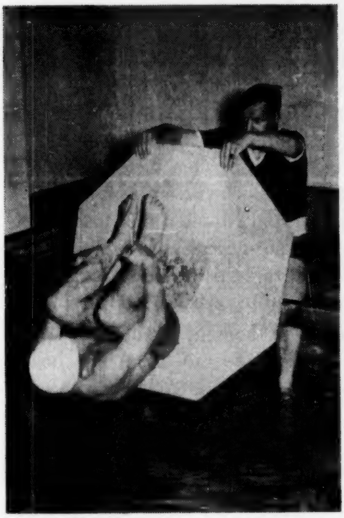


FIGURE III. A subject falling into the water.

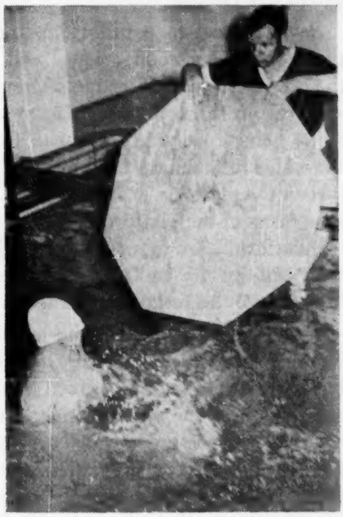


FIGURE IV. A subject's head appearing above the water after immersion.

listed by Lawther (6) reads as follows: "Auditory cues, when occurring close to the athlete, are responded to more quickly than visual cues."

Organization of Data. The time scores of each subject were recorded and then averaged into two separate mean scores, one representing the eyes open test and the other the eyes blindfolded test.

The subjects were grouped in two ways: (1) on the basis of the kind of deafness and (2) on the basis of ability to maintain equilibrium. In grouping on the basis of deafness, five groups were formed: born deaf, meningitis, other causes, undetermined, and normal. The records of their causes of deafness were obtained from the files in the Office of the Registrar at Gallaudet College. These groups were selected on the basis of its being customary to group the deaf and hard of hearing into two main categories, namely, congenitally deaf and adventitiously deaf. Congenital deafness is present at the time of birth and adventitious deafness occurs after birth through illness or accident. In the case of the adventitiously deafened, it seemed advisable to subdivide into two groups. One group was composed of meningitis cases and the other of all other causes of deafness such as mastoiditis, measles, mumps, etc. The reason for separating meningitis from the other causes was that meningitis is thought to be the foremost cause of sufficient labyrinthal injury to induce deafness or loss of equilibrium or both. There were some subjects whose causes of deafness were unknown so another group was established to include these undetermined cases. Finally, for the sake of comparison, a fifth group was formed, composed of normal hearing subjects.

With respect to equilibrium, the subjects were placed into two groups, those exhibiting good balance and those lacking the same. Classification of the subjects was subjective in that it was accomplished by observing the subjects on the rotating platform and their ability to maintain equilibrium while being turned around with the eyes blindfolded. Most of the subjects were able to maintain their equilibrium fairly well even when the investigator turned the platform around rapidly, alternated the clockwise or counterclockwise turns of the platform, or alternately stopped and resumed the rotation of the platform. A minority of the subjects were not so successful in maintaining their equilibrium and the investigator was forced to turn the platform slowly or to pause the rotation in order to restrain the subjects from toppling over.

It should be noted that the meningitis and poor balance groups were not identical. The poor balance group included most of the subjects whose hearing was rendered nonfunctional by the disease of meningitis and four subjects obtained from the other groups of the first classification, two born deaf, one measles, and one unknown. These last four subjects might well have had vestibular labyrinths accompanying their deafness. Six of the 30 meningitis cases showed surprisingly good balance while sitting on the rotating platform, and for that reason they were classified in the good balance group. The disease might have left their labyrinths intact while causing deafness.

It is said that it is possible for a person to have nonfunctioning labyrinths while his hearing is normal. The normal hearing subjects, however, all demonstrated good balance and were included in the good balance group.

Statistical Treatment of Data. First, the reliabilities of the data for the eyes open and eyes blindfolded tests were determined. Thirty subjects chosen at random from the different groups repeated both tests on different days ranging from a few days to several weeks under conditions as nearly identical as possible. Coefficients of correlation between the test and re-test performances were computed. The t ratio was used to determine the statistical significance of the mean differences between the performances. Second, the differences between the means of the different groups were tested against the null hypothesis based on the "students" distribution to determine their significance. The means, along with the variances and the standard deviations, were calculated from the average time scores on the record sheets. The average time scores were actually the averages of 20 trials which each subject was requested to take in each of the two tests. The F ratio for variances was also computed for items demonstrating significant mean differences at the .05 level of confidence.

Findings

Reliability of Data. The reliability coefficients by the test re-test method were found to be .74 for the eyes open test and .85 for the eyes blindfolded test. As for the significance of the differences between the means of the test and re-test performances, a t ratio of 5.99, significant at the 1 percent level of confidence, was found on the eyes open test and a t ratio of only 1.79 was for the eyes blindfolded test. While some improvement was shown in both tests, only the improvement in the eyes open test was found to be statistically significant, indicating that learning probably occurred in that test.

Equivalence of Groups. In Table 1 are listed the average ages, heights, and weights of the different groups in both classifications. The groups were ap-

TABLE 1.—NUMBER AND AVERAGE AGE, HEIGHT, AND WEIGHT OF GROUPS PARTICIPATING IN THE ORIENTATION TESTS

Classification	Number	Age	Height	Weight
Congenital (Born Deaf)	28	21	5′10″	159
Adventitious (Meningitis)	30	21	5′ 9″	159
(Other Causes)	30	21	5′ 9″	158
Undetermined	17	21	5′10″	163
Normal	23	22	5′10″	157
Good Balance	100	21	5′10″	158
Poor Balance	28	21	5′10″	160

proximately of the same age and body build. This was in accordance with the efforts of the investigator to equate the groups on these factors.

Differences in Orientation with Eyes Open and Blindfolded. The means, variances, and standard deviations are shown in Table 2. In Table 3 is presented the significance of the mean differences for the various groups between the eyes open and eyes blindfolded tests. It is interesting that the meningitis group in the first classification and the poor balance group in the second classification were the only groups that failed to show any decrease in the eyes blindfolded test over the eyes open test. However, the obtained t ratios of 1.27

TABLE 2.—MEANS, VARIANCES, AND STANDARD DEVIATIONS OF GROUPS FOR THE EYES OPEN AND EYES BLINDFOLDED TESTS

		Eyes Open Test			Eyes Blindfolded Test		
Classification	Number	Mean	Variance	Standard Deviation	Mean	Variance	Standard Deviation
Congenital (Born Deaf)	28	1.74	0.22	0.47	1.62	0.16	0.40
Adventitious (Meningitis)	30	2.08	0.23	0.48	2.26	0.39	0.63
(Other Causes)	30	1.77	.019	0.43	1.68	0.13	0.36
Undetermined	17	1.73	0.16	0.40	1.56	0.08	0.29
Normal	23	1.58	0.10	0.32	1.52	0.12	0.34
Good Balance	100	1.71	0.16	0.40	1.60	0.12	0.34
Poor Balance	28	2.12	0.24	0.49	2.32	0.40	0.63

TABLE 3.—SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS FOR THE EYES OPEN AND EYES BLINDFOLDED TESTS

Classification	Degrees of Freedom	Mean Difference	t Ratio	F Ratio
Congenital (Born Deaf)	27	0.12	1.03	1.34
Adventitious (Meningitis)	29	0.18	1.27	1.70
(Other Causes)	29	0.09	0.88	1.47
Undetermined	16	0.17	1.42	1.97
Normal	22	0.06	0.61	1.16
Good Balance	99	0.11	2.12[a]	1.39
Poor Balance	27	0.20	1.32	1.65

[a] Significant at the 5 percent level of confidence.

for the meningitis group and 1.32 for the poor balance group were significant only at the .20 level of confidence and not at the .05 or .01 levels set for rejection of the null hypothesis. As for the remaining five groups which made a better performance in the eyes blindfolded test than in the eyes open test, the differences were also found not to be statistically significant within the 1 or 5 percent level of confidence. The t ratio of 2.12 for the good balance group, however, was significant at the 5 percent level of confidence. No F ratios were significant at the 5 percent level of confidence.

Although not statistically significant, it seems important that the meningitis and poor balance groups apparently experienced more difficulty in the eyes blindfolded test than in the eyes open test. This was made more convincing by the fact that the other groups reacted conversely, showing better performance in the eyes blindfolded test than in the eyes open test.

Differences between Groups with Eyes Open. The differences between the mean performances for the various groups for the eyes open test are presented in Table 4. The meningitis group was the only one in the first classification that differed significantly from all the other groups. The t ratios of 2.58 between the meningitis and other causes groups and of 2.50 between the meningitis and undetermined groups were statistically significant at the 5

TABLE 4.—SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF GROUPS IN THE EYES OPEN TEST

Group	Degrees of Freedom	Mean Difference	t Ratio	F Ratio
Born Deaf vs. Meningitis	56	0.34	2.83[a]	1.06
Born Deaf vs. Other Causes	56	0.03	0.25	1.18
Born Deaf vs. Undetermined	43	0.01	0.07	1.37
Born Deaf vs. Normal	49	0.16	1.45	2.19
Meningitis vs. Other Causes	58	0.31	2.58[b]	1.25
Meningitis vs. Undetermined	45	0.35	2.50[b]	1.45
Meningitis vs. Normal	51	0.50	4.17[a]	2.32
Other Causes vs. Undetermined	45	0.04	0.31	1.17
Other Causes vs. Normal	51	0.19	1.73	1.86
Undetermined vs. Normal	38	0.15	1.36	1.60
Good Balance vs. Poor Balance	0.41	126	4.56[a]	1.49

[a] Significant at the 1 percent level of confidence.
[b] Significant at the 5 percent level of confidence.

percent level of confidence. The t ratio for the meningitis group and two other groups, born deaf and normal, were significant at the 1 percent level. In the second classification, the difference between the means of the good balance and poor balance groups was statistically significant within the 1 percent level of confidence. None of the F ratios was found to be significant at the 5 percent level of confidence.

The results of the eyes open test indicate that even with vision unobstructed the meningitis and poor balance groups had more trouble retaining their sense of orientation under water and needed a longer time for finding their way to the surface than the other groups.

Differences between Groups with Eyes Blindfolded. From Table 5, it can be seen that in the eyes blindfolded test none of the groups in the first classification differed significantly from each other, except the meningitis group. This group differed significantly from all of the other groups at the 1 percent level of confidence. Likewise the mean of the poor balance group was significantly different from that of the good balance group well beyond the 1 percent level. The F ratios between the meningitis group and each of the four remaining groups and between the good balance and poor balance groups were statistically significant within the 5 percent level of confidence.

Variation in times for the meningitis and poor balance groups was greater

TABLE 5.—SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF GROUPS IN THE EYES BLINDFOLDED TEST

Group	Degrees of Freedom	Mean Difference	t Ratio	F Ratio
Born Deaf vs. Meningitis	56	0.64	4.57[a]	2.42[c]
Born Deaf vs. Other Causes	56	0.06	0.60	1.29
Born Deaf vs. Undetermined	43	0.06	0.54	2.00
Born Deaf vs. Normal	49	0.10	0.91	1.41
Meningitis vs. Other Causes	58	0.58	4.46[a]	3.11[b]
Meningitis vs. Undetermined	45	0.70	4.38[a]	4.86[b]
Meningitis vs. Normal	51	0.74	4.93[a]	3.40[b]
Other Causes vs. Undetermined	45	0.12	1.20	1.56
Other Causes vs. Normal	51	0.16	1.60	1.09
Undetermined vs. Normal	38	0.04	0.40	1.43
Good Balance vs. Poor Balance	126	0.72	8.00[a]	3.44[b]

[a] Significant at the 1 percent level of confidence.
[b] Significant at the 2 percent level of confidence.
[c] Significant at the 4 percent level of confidence.

than for the other groups as was shown by the significant F ratios; hence, the t test for difference between means was not conclusive because of the possible effect of variance in times on the means.

Discussion

Concerning the test re-test performances of the 30 randomly picked subjects, the results showed that the subjects made a significant improvement in the eyes open test and an insignificant improvement in the eyes blindfolded test. However, it is difficult to determine that learning occurred in the eyes open test and not in the eyes blindfolded test due to the procedure of administering the tests. It may be possible that the reliability of the eyes open test was relatively poor because it was always the first test administered in the procedure or, on the other hand, that the reliability of the eyes blindfolded test was good because it was always the second test administered in the procedure. Had the order of conducting the two tests been reversed, the results might have been reversed.

The results of the investigation as applied to the effect of submersion upon the sense of orientation of deaf swimmers seem to indicate that in both the eyes open test and eyes blindfolded test given to the deaf and hard of hearing subjects as well as to the normal subjects, the meningitis and poor balance groups had considerably more difficulty in coming to the surface than the other groups. The results also show that unlike the other groups they performed more poorly in the eyes blindfolded test than in the eyes open test although at a statistically insignificant level. During the course of the experiment it was the meningitis cases or the subjects with poor balance who hindered the study to some extent, especially in the eyes blindfolded test, by repeatedly letting go the hold on the legs, touching the bottom of the pool, or pulling off the blindfold before appearing above the surface. They touched the bottom by pure accident when they swam down thinking they were ascending and through touch they quickly and easily found the surface. Sometimes they lost their bearings under water and swam toward one of the sides and upon touching the side, they immediately shoved off in the opposite direction, thinking they had touched the bottom and were going upward, when actually they were swimming toward the other side. When they reached the other side and found themselves still under water, they became bewildered and tore off the blindfold in order to depend on vision for orientation. One of them stayed under water for fifteen seconds floundering and groping around for the surface and failing to find it, gave up and took off the blindfold.

Fear of losing orientation under water was very much evident among some of the meningitis or poor balance cases, especially in the eyes blindfolded test. This was characterized by their bodies becoming tense while being turned around on the rotating platform preparatory to being thrown into the water and by their swallowing some water while pawing and clawing frantically to rise to the surface. Four subjects had to be excused from the experiment when

they showed lack of confidence in themselves by repeatedly letting go the hold on their legs even before hitting the water. They were excused because they could prevent themselves from becoming completely submerged by allowing a greater area of body surface to spread over the water, which enabled them to make faster time scores. There were two other subjects who refused to try the eyes blindfolded test after completing the eyes open test so their initial scores were not included in the tabulations.

Due to the criteria for the orientation tests, the trials in which the subjects failed to follow the directions properly had to be cancelled. It was at first feared that the criteria might handicap the study and limit the findings; nevertheless, at the conclusion of the experiment the results indicated that the meningitis and poor balance groups apparently differed from the other groups. If the cancelled trials or the withdrawals had been included in the tabulations, the differences between the groups might have been greater.

While it is obvious that the meningitis and poor balance groups experienced more difficulty in the orientation tests than the other groups, it must be pointed out that the experience of losing orientation under water did not occur to every subject in the meningitis group or in the poor balance group and also that no subject lost direction every time he was submerged. This was reflected by the difference in variances between the meningitis and poor balance groups and the other groups as indicated by the F ratios (Table 5). It is possible that the deficiency of the vestibular sense may vary in different subjects just as the deficiency of vision or hearing may vary or it might be due to the fact that these individuals are more alert to extrinsic cues such as body buoyancy.

Also, chance could easily have caused this large variance in the meningitis and poor balance groups. If they chanced to move upward toward the surface they would have very excellent times; if they went to the bottom, then a poor time. The fact that the subjects in the other groups always came to the surface would account for greater homogeneity and less variance. However, the study in no way attempted to investigate individual differences within a group but rather the relationship between groups that might be expected to exist. For this reason it is difficult to determine to what extent the loss of orientation was incurred as a result of vestibular deficiency.

It is also possible that other factors may have influenced the results of the study, causing the meningitis and poor balance groups to be noticeably poorer. The fatigue element from repeated submersions could have affected the findings. Swallowing water while trying to rise to the surface because of fear might have taken too much energy out of the subjects and slowed their times. The oft-cancelled trials of the meningitis and poor balance groups, of course, did not help them perform better during the tests. Another factor causing the poor showing may have been the establishment of only the minimum standard level of swimming ability; it was impossible to equate the subjects so the

groups would have equal ability. The meningitis and poor balance cases because of their disability perhaps had had less opportunity or desire to learn to swim as compared to the other groups and this might have caused their relatively poor showing.

Summary

1. The reliabilities of the eyes open and eyes blindfolded tests were .74 and .85 respectively; these orientation tests were repeated by 30 subjects picked from the different groups at random.

2. All the groups, with the exception of the meningitis group in the first classification and the poor balance group in the second classification, showed an increase in the eyes blindfolded test over the eyes open test. The meningitis and poor balance groups fared less satisfactorily in the eyes blindfolded test. Of all the differences between the means of the two tests only the increase made by the good balance group was found to be significant within the 5 percent level of confidence.

3. In the eyes open test the meningitis group was the only one in the first classification that differed significantly from all the other groups; the other causes and undetermined groups differed from the meningitis group at the 5 percent level of confidence and the born deaf and normal groups at the 1 percent level of confidence. In the second classification the mean of the poor balance group was significantly different from that of the good balance group within the 1 percent level of confidence.

4. In the eyes blindfolded test none of the groups in the first classification differed significantly from each other except the meningitis group. This group differed significantly from all of the other groups at the 1 percent level of confidence. The difference between the means of the good balance and poor balance groups in the second classification was significant well beyond the 1 percent level of confidence.

5. In the two tests, the differences among the means of the born deaf, other causes, undetermined, and normal groups were not statistically significant at either the 1 percent or the 5 percent level of confidence.

Conclusion

These findings, although indicating the inability of the meningitis or poor balance groups to orient themselves under water, do not necessarily mean that they should be prohibited from swimming. It does indicate, however, that special attention must be given individuals suffering from deafness due to meningitis both during the instructional phase and when they are swimming for recreation.

References

1. Best, Charles Herbert, and Taylor, Norman Burke. *The Physiological Basis of Medical Practice.* Baltimore: Williams and Wilkins Company, 1950.
2. Camis, Mario. *The Physiology of the Vestibular Apparatus.* Oxford: Clarendon Press, 1930.

3. DAVIS, HALLOWELL. *Hearing and Deafness.* New York: Murray Hill Books, 1947.
4. EDWARDS, AUSTIN S. "Measurements of Static Ataxia." *American Journal of Psychology* **55**: 171-88; April 1942.
5. JAMES, WILLIAM. "The Sense of Dizziness in Deaf-Mutes." *American Journal of Otology* **4**: 239-54; Oct. 1882.
6. LAWTHER, JOHN D. *Psychology of Coaching.* New York: Prentice-Hall, 1951.
7. MORSH, JOSEPH EUGENE. "A Comparative Study of Deaf and Hearing Students." *American Annals of the Deaf* **82**: 223-33; May 1937.
8. OBERSTEINER, HEINRICH. "The Maintenance of the Equilibrium as a Function of the Central Nervous System." *American Naturalist* **33**: 313-29; April 1899.
9. PRENTISS, C. W. "The Otocyst of Decapod Crustacea: Its Structure, Development, and Functions." *Bulletin of the Museum of Comparative Zoology* **36**: 167-251; July 1901.
10. STREETER, GEORGE L. "Some Experiments on the Developing Ear Vesicles of the Tadpole with Relation to Equilibration." *Journal of Experimental Zoology* **3**: 543-58; Dec. 1906.

(Submitted 7/9/58)

The Relationship of Movement Time and Reaction Time from Childhood to Senility[1]

WILLIAM R. PIERSON
Research Center
College of Osteopathic Physicians and Surgeons
Los Angeles, California

Abstract

Four hundred male subjects between the ages of 8 and 83 were measured for reaction time and movement time. A statistical analysis of the data permits the conclusion that reaction time and movement time are significantly related.

THE LACK OF a statistically significant correlation between simple reaction time (RT) and movement time (MT) has been cited (6) as the basis for the conclusion that these attributes are independent and unrelated. However, it is notable that when subjects other than male college students are used, results indicate that there may be a positive correlation between RT and MT (8, 18). It was the purpose of the present study to investigate the extent of these relationships for male subjects from childhood to senility. A secondary consideration was the comparison of the results obtained for college-age students with those obtained for the entire age range.

In 1952 Henry (6) reported a study in which he found no correlation between simple reaction time and the duration of a discrete movement. He concluded that the two functions must, therefore, be considered as independent and unrelated. Slater-Hammel (14) likewise found no correlation between RT and MT and in addition indicated that the method of movement termination has no pronounced effect upon the relationship. He did not agree with Henry about the significance of the fact that RT and MT appear to be independent and uncorrelated. The lack of correlation means only that the possibility of a "slow reactor" having a fast MT is as great as it is for a "fast reactor." The results of Slater-Hammel's study were interpreted as simply indicating that measurement of RT cannot readily be used to predict speed of movement. As adjuncts of other psychomotor studies, Pierson (11), Sills (13), and Henry's students (4, 9) demonstrated that RT and MT are not correlated. In all the above studies the subjects were male college students. However, Hipple (8) reported a "possibly significant" correlation from the measurement of 12- to 14-year-old white male subjects and Youngen (18), a low but statistically significant correlation for female college students.

[1]The data for this study were collected in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Teacher Education, School of Advanced Graduate Studies, Michigan State University. The author also wishes to express his appreciation for the assistance of H. J. Montoye and the Department of Health, Physical Education, and Recreation.

Procedure

The apparatus of Pierson (11) was modified to include an additional chronoscope[2] in order to terminate MT by interruption of a light beam to a photoelectric cell. The response unit of the apparatus (Figure I) was mounted on a table approximately 32 in. from the floor and the subject seated before it. The subject was then placed so that full extension of the dominant arm would put the fingertips about one inch beyond the light beam. The subject was instructed to release a microswitch and extend his hand through this beam in response to a neon lamp stimulus. Activation of a chronoscope and the stimulus lamp was simultaneous, and in responding to the stimulus the subject released a microswitch which stopped the chronoscope. Release of this microswitch also activated a second chronoscope, which in turn was stopped by the interruption of the light beam. Reaction time was then read from the first chronoscope and movement time from the second. The lift of the finger from the microswitch was the only movement involved in the reaction time, and MT was measured over a distance of 11 in.

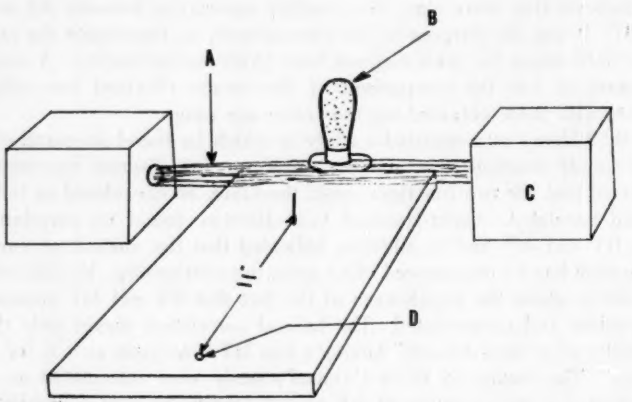


FIGURE I. Apparatus used to measure reaction and movement times: A—light beam, B—stimulus lamp, C—photo-electric cell, D—microswitch (dp/dt).

A preliminary study of 50 subjects indicated that for the apparatus and response required in the present study optimum measures could be expected between trials 15 and 30. Therefore, 30 trials were recorded for each subject, and the mean of the last 15 was considered the representative score.

[2]Made available through the courtesy of S. Howard Bartley and the Department of Psychology, Michigan State University.

The data were programed through IBM machines. The correlation (Pearson product-moment) of RT and MT for the 400 subjects was calculated, and, since the age range of college students is generally from 19 to 25, a correlation for the 40 subjects in that age range also was computed. The sample was then stratified so that there were 20 subjects in each of 20 age groups, and a correlation of each of these groups was computed. No correlation coefficient was considered statistically significant unless the probability of the chance occurrence of such a statistic was 1 percent or less.

Results

Statistical analysis of the data for the 400 subjects resulted in a significant correlation coefficient ($r = 0.56$). The estimate of r for the total population, based on a z-transformation and using one standard error as the limit, is from 0.54 to 0.58. When allowances for the effects of age are made by referring all scores to the regression curve established by Pierson and Montoye (10), the correlation between RT and MT remains significant ($r = 0.31$). The coefficient obtained for the college-age subjects was not statistically significant ($r = 0.35$).

The statistical techniques for other than linear data are the correlation ratio (*eta*) (1,16) and the index of correlation (*rho*) (3,5,17). When computed for the same data, *eta* is equal to or greater than the correlation coefficient *r* (1). Although Pierson and Montoye (10) have demonstrated that RT and MT are curvilinear functions of age, for the present study analyses using correlation techniques other than linear would be redundant, since the magnitude of the r exceeds the requirements for statistical significance established before the collection of the data. The results of the correlation analyses for the 20 age groups are presented in Figure II.

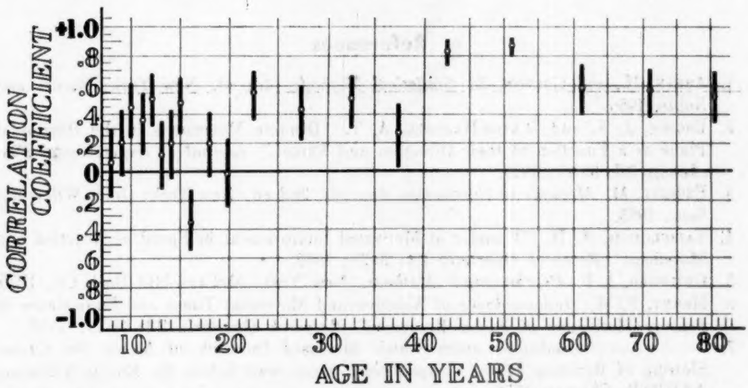


FIGURE II. The correlation of movement time and reaction time by age group. The heavy vertical lines indicate one standard error of the correlation as computed by a z-transformation.

Discussion

Although the results for the college-age subjects substantiated the findings of previous studies, they are not representative of the age range 8 to 83. From the studies of Slater-Hammel (14), Henry (7), and Brown and Slater-Hammel (2), it would appear that the reason for this is not to be found in differences of movement termination or in the "set" of the individual. Possible explanations for the difference in results obtained for the college-age subjects and the total sample may be in sample variance or in some inherent difference of certain age groups. However, it was not within the scope of the present study to determine the reasons for differences between age groups.

On the basis of the data presented, the following inferences appear justified:

1. Considerable chance for error is incurred when conclusions concerning the adult male population are drawn from a sample of male college students.
2. The relationship of RT and MT may be a function of maturity or incidence of employment, since the correlation for subjects over 21 years of age ($r = 0.63$) differs from that for subjects between ages 8 and 21 ($r = 0.50$).

Summary and Conclusions

Four hundred male subjects, age range 8 to 83, were measured for reaction time and movement time by a fractioning process. The correlation of these variables was computed for the 400 subjects as well as for certain age groups, and an analysis of these computations permits the conclusion that for males between the ages of 8 and 83 there is a statistically significant correlation between reaction time and movement time as measured in this study.

References

1. Arkin, H., and Colton, R. *Statistical Methods.* 4th ed. New York: Barnes and Noble, 1956.
2. Brown, J. S., and Slater-Hammel, A. T. "Discrete Movements in the Horizontal Plane as a Function of their Direction and Extent." *Journal of Experimental Psychology* **38**: 84-95; 1949.
3. Ezekiel, M. *Methods of Correlation Analysis,* 2nd ed. New York: John Wiley and Sons, 1953.
4. Fairclough, R. H. "Transfer of Motivated Improvement in Speed of Reaction and Movement.' *Research Quarterly* **23**: 20-27; 1952.
5. Guilford, J. P. *Psychosomatic Methods.* New York: McGraw-Hill Book Co., 1936.
6. Henry, F. M. "Independence of Reaction and Movement Times and Equivalence of Sensory Motivators of Faster Response." *Research Quarterly* **23**: 43-53; 1952.
7. ————. "Conditions under which Increased Intensity of Motor Set Causes Slowing of Reaction Time." Unpublished paper read before the Research Section AAHPER, Chicago, 1956.
8. Hipple, J. E. "Racial Differences in the Influence of Motivation on Muscular Ten sion, Reaction Time, and Speed of Movement." *Research Quarterly* **25**: 297-306; 1954.

9. Howell, M. L. "Influence of Emotional Tension on Speed of Reaction Time and Movement." *Research Quarterly* **21**: 22-32; 1953.
10. McNemar, Q. "Sampling in Psychological Research." *Psychological Bulletin* **37**: 331-365; 1940.
11. Pierson, W. R. "Comparison of Fencers and Nonfencers by Psychomotor, Space Perception, and Anthropometric Measures." *Research Quarterly* **27**: 90-96; 1956.
12. Pierson, W. R., and Montoye, H. J. "Movement Time, Aeaction Time and Age." *Journal of Gerontology* **13**: 418-21; 1958.
13. Sills, F. D. "An Analysis of Speed of Movement." *AAHPER Convention Proceedings, New York, 1954.* Washington, D. C.; the Association, 1954.
14. Slater-Hammel, A. T. "Reaction Time and Speed of Movement." *Perceptual and Motor Skills, Research Exchange* **4**: 109-13; 1952.
15. Teichner, W. H. "Recent Studies of Simple Reaction Time." *Psychological Bulletin* **51**: 128-49; 1954.
16. Walker, H. M., and Lev, J. *Statistical Inference.* New York: Henry Holt and Co., 1953.
17. Waugh, A. E. *Elements of Statistical Method.* New York: McGraw-Hill Book Co. 1943.
18. Youngen, L. J. *A Comparison of Reaction Time and Movement Time Measures of Women Athletes and Nonathletes.* Unpublished paper read before the Research Section, Midwest Association of Health, Physical Education, and Recreation, Detroit, 1957.

(Submitted 8/26/58)

An Analysis of the Aerodynamics of Pitched Baseballs[1]

CARL SELIN
University of California
Riverside, California

Abstract

The purpose of this study was to analyze the trajectories of pitched baseballs and to determine the causes of particular types of trajectories. More than 200 pitches, made by 14 pitchers of baseball teams in the Western (Big Ten) Conference, were photographed by means of two high-speed cameras. The films upon which were recorded the complete trajectories of pitches were analyzed in terms of velocity and rate of rotation of the ball. From each category of pitch, six selected pitches were further analyzed in terms of direction of rotation, axis of rotation, vertical deviation, horizontal deviation, vertical forces, and horizontal forces.

FOR MANY YEARS there has been a controversy concerning the trajectories of various types of pitched baseballs (1, 2, 3, 6, 7, 8, 12, 15, 17, 18, 19, 20). Several studies have resulted in controversial claims about the hop of a fast ball (14, 16), the flutter of a knuckle ball (13), and the break of curve and drop balls (4, 8, 10, 21).

The trajectory of a pitched baseball is determined by the interaction of a number of forces. These forces include (1) the initial force exerted on the ball by the pitcher, (2) the force of gravity, and (3) the forces that are determined by the velocity of the ball relative to the air and by the rate and the direction of rotation of the ball (3, 12).

Outstanding pitchers today use a variety of pitches. These pitches include the fast ball, the curve ball, the slider, the slow ball, the knuckle ball, the fingernail ball, the screw ball, and the fork ball (5, 22).

Definition of Terms

Since baseball terminology is lacking in preciseness, it was necessary for the purpose of this study to formulate some definitions of terms.

Vertical angle is defined as the angle between a vertical line drawn through the center of the ball and the axis of rotation of the ball. As viewed from the pitching rubber, the vertical angle is measured in a clockwise direction from the vertical line.

Horizontal angle is defined as the angle between the direction of the pitch and the axis of rotation of the ball. As viewed from above, the horizontal angle is measured in a clockwise direction from the direction of pitch to the axis of rotation.

Vertical deviation is defined as the vertical distance, measured directly above the cen-

[1]This study was completed in partial fulfillment of the requirements for the degree of doctor of philosophy in the Division of Physical Education, Graduate College, State University of Iowa, 1957, under the direction of Louis E. Alley, Division of Physical Education, and Elmer C. Lundquist, Department of Aeronautics, College of Engineering.

ter of the home plate, between the trajectory of a pitched ball and the trajectory of a freely falling object projected at the same angle as the pitched ball and upon which no forces other than gravity act. Upward deviation is regarded as positive, and downward deviation is regarded as negative.

Horizontal deviation is defined as the horizontal distance, measured from the center of the home plate in a direction perpendicular to the line extending from the pitching rubber to the center of the home plate, between the trajectory of a pitched baseball and a straight line extended in the direction of the pitch. Deviation to the right (as a pitch is viewed by the pitcher) is regarded as negative.

Rotations of the ball are described as *clockwise* and *counterclockwise* in terms of a clock facing the viewer who observes the ball traveling from his right to his left.

Procedure

More than 200 pitches made by 14 pitchers from the Big Ten Conference were photographed by two high-speed cameras, one of which operated at 64 frames a second and the other at 128 frames a second. Each pitcher made about 20 pitches, each of which was photographed from the side and from above simultaneously. These pitches included several fast balls, several curve balls, and other types of pitches in the particular pitcher's repertoire.

One line was placed on the ground, and another four feet above the ground on the background wall. These lines extended from the pitching rubber to the home plate. On both lines, markers indicated ten-foot intervals along the path of the pitch. These markings were used to determine the exact point of departure of the pitched ball from a normal trajectory. In the vertical plane the trajectory of each pitched ball was recorded by the ground-level camera, and in the horizontal plane the trajectory was recorded by the camera directly above the line of flight of the ball.

The films upon which were recorded the complete trajectories of pitches were analyzed in terms of velocity and rate of rotation of the ball. The time required for the pitch to complete its trajectory was obtained from the photographs of two synchronized chronoscopes, one of which was placed adjacent to the pitcher and the other adjacent to the plate. The length of the trajectory was measured from the point of release to the point on the trajectory that was directly above a line drawn through the center of the plate and parallel to the rubber on the pitcher's mound. The mean velocity of the pitch was determined by dividing the length of the trajectory by the observed time-interval.

One-half of the surface of the ball was painted black to facilitate, for each pitch, the observation of the rate of rotation of the ball. The number of revolutions for each pitch was observed from enlargements of the film. To determine the rate of rotation, this number was then divided by the observed time-interval for the pitch.

In order to compare the deviations of the pitched balls thrown by two left-handed pitchers with the balls thrown by the 12 right-handed pitchers used in this study, the directions of rotation, the angles of the axis of rotation, and the trajectories were translated so that the pitched balls appeared to have been thrown by a right-handed pitcher.

Six selected pitches were chosen from each of the following categories of pitches: fast balls, curve balls, knuckle balls, and change-ups. The pitches selected from each category were (1) the pitch with the lowest velocity, (2) the pitch with the velocity nearest the average velocity for all such pitches, (3) the pitch with the highest velocity, (4) the pitch with the lowest rate of rotation, (5) the pitch with the rate of rotation nearest the average rate of rotation for all such pitches, and (6) the pitch with the highest rate of rotation. Only four sliders and two side-arm curves were thrown, and all these pitches were selected for analysis.

In both the vertical and horizontal planes, points on the trajectory of each of the 30 selected pitches were plotted at ten-foot intervals, and the velocity and the rate of rotation for each ten-foot segment were determined. Each of the 30 selected pitches was further analyzed in terms of direction of rotation, angle of axis of rotation, vertical deviation, horizontal deviation, vertical forces, and horizontal forces.

For each of the 30 selected pitches, three graphs were plotted: (1) a graph in which distance was plotted against time, (2) a graph in which deviation in the vertical plane was plotted against time and in which was plotted the trajectory of an object with no deflecting forces other than gravity acting upon it (Figure I), and (3) a graph in which the horizontal deviation from

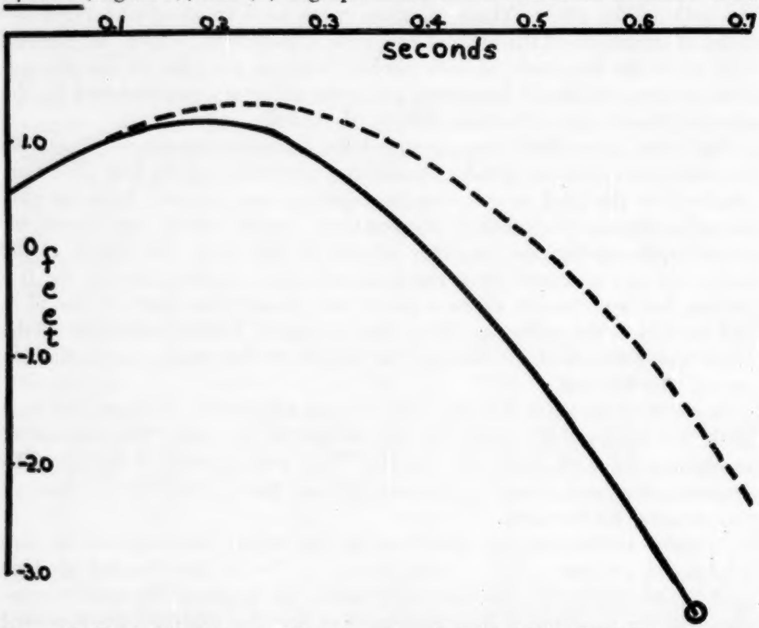


FIGURE I. Vertical Deviation vs. Time for Curve Ball (Ia.2-10). Solid line represents trajectory of pitched ball; dotted line represents trajectory of freely falling object; circle represents home plate.

a straight line that was extended in the initial direction of the pitch was plotted against time.

The magnitude of the forces other than gravity that, in the vertical plane, acted upon the ball was determined by first solving for acceleration a in the formula $S = \frac{1}{2}at^2$, where S is the amount of vertical deviation, and t is the time required for the pitch (Figure I). Since the mass m was known and since a was found, the force F was computed from the formula $F = ma$. The magnitude of the forces that, in the horizontal plane, acted upon the ball was determined as was the magnitude of the forces that acted in the vertical plane except that S in the preceding formula becomes the amount of deviation in the horizontal plane.

Findings

Velocity. Each of the graphs in which distance was plotted against time shows a straight line, which fact indicates that, within the limits of the accuracy of the measurements obtained, there was throughout each pitch no change in the velocity of the ball. For each type of pitch the range in velocities and the mean velocity in feet per second are shown in Table 1.

TABLE 1.—RANGES AND MEANS OF VELOCITIES OF 30 SELECTED PITCHES (fps)

Types of Pitches	Ranges	Means
Fast ball	87.5 — 121	104
Slider	92 — 102	98
Curve ball	74 — 111	91
Sidearm-curve ball	90 — 92	91
Knuckle ball	75 — 91	84
Change-up	62.5 — 90	79

Rates of Rotation. Throughout each pitch the rate of rotation remained constant. For each classification of pitch the range in rates of rotation and the mean rates of rotation in revolutions per second are shown in Table 2.

TABLE 2.—RANGES AND MEANS OF RATES OF ROTATION OF 30 SELECTED PITCHES (rps)

Types of Pitches	Ranges	Means
Sidearm-curve ball	29.5 — 32.8	31.2
Curve ball	19.2 — 38.5	29.9
Slider	24 — 30.2	28
Fast ball	20.8 — 31.7	26.6
Change-up	18.7 — 27.5	24.5
Knuckle ball	2.9 — 9.2	5.3

Axis of Rotation. For each type of pitch the range and the mean of the vertical and horizontal angles in degrees are shown in Table 3.

TABLE 3.—RANGES AND MEANS OF VERTICAL AND HORIZONTAL ANGLES OF 30 SELECTED PITCHES (degrees)

Types of Pitches	*Ranges* Vertical Angles	Horizontal Angles	*Means* Vertical Angles	Horizontal Angles
Sidearm-curve ball	150	30– 45	150	38
Curve ball	120–160	30–160	141	86
Fast ball	100–170	90–160	131	105
Knuckle ball	125–135	30–135	131	73
Slider	120–135	45–150	125	116
Change-up	90–135	15–160	116	89

Direction of Rotation. All the fast balls, two change-up pitches, and four knuckle balls rotated in a clockwise direction. The curve balls, the sliders, and the sidearm-curve balls rotated in a counterclockwise direction.

Vertical Deviation. All the pitches that rotated in a clockwise direction showed positive vertical deviation, and all the pitchs that rotated in a counterclockwise direction showed either negative vertical deviation or approximated zero deviation. For each type of pitch the range of vertical deviation in feet is shown in Table 4.

TABLE 4.—RANGES OF VERTICAL DEVIATIONS OF 30 SELECTED PITCHES (ft.)

Types of Pitches	Ranges	Types of Pitches	Ranges
Fast ball	.3 to 2.7	Change-up[a]	−2.0 to 2.4
Curve ball	0 to −1.8	Knuckle ball[a]	−2.0 to 2.25
Slider	0 to −2.0	Sidearm-curve ball	0 to −1.6

[a]Some pitches rotated in a clockwise direction, and other pitches rotated in a counterclockwise direction.

The trajectories of three of the curve balls showed very little vertical deviation at the time the balls crossed the plate. In these three pitches, and in several pitches in the categories of slider and sidearm-curve balls, the trajectory of the pitch and the trajectory of a freely falling body diverged during the early stages of the pitch and converged during the later stages of the pitch. The graph of each of these pitches shows the effects of vertical forces other than gravity, which during the early stages of the pitch acted in an upward direction on the ball, and which during the later stages of the pitch acted in a downward direction so that at the instant the ball crossed the

plate no deviation was apparent. This divergence and subsequent convergence of trajectories probably resulted from changes in the vertical angle and the horizontal angle, which changes resulted from precession. To a batter each of these diverging and converging trajectories would appear to have a marked vertical deviation or drop.

Horizontal Deviation. Each of the pitches that rotated in a clockwise direction, with the exception of two knuckle balls, displayed negative horizontal deviation. All the pitches that rotated in a counterclockwise direction displayed positive horizontal deviation. For each type of pitch the range of horizontal deviations in feet is shown in Table 5.

TABLE 5.—RANGES OF HORIZONTAL DEVIATIONS OF 30 SELECTED PITCHES (ft.)

Types of Pitches	Range	Types of Pitches	Range
Fast ball	—.3 to —2.1	Change-up[a]	—.6 to 1.5
Curve ball	.1 to 1.2	Knuckle ball[a]	—.8 to 2.5
Slider	.4 to .8	Sidearm-curve ball	.5 to .8

[a]Some pitches rotated in a clockwise direction, and other pitches rotated in a counterclockwise direction.

The directions of the horizontal deviations of the change-up pitches were inconsistent: the deviations of two change-up pitches were negative, and the deviations of the others were positive. The negative deviations of two of the pitches resulted from the clockwise direction of rotation of the ball. The other four change-up pitches rotated in a counterclockwise direction and behaved essentially as slow curve balls.

In all the categories of pitches, the greatest deviation in the horizontal plane was displayed by one of the knuckle-ball pitches. One pitch had a positive deviation of 2.5 feet. Several inconsistencies in the directions of the deviations occurred in the knuckle-ball category. Of the knuckle balls that rotated in a counterclockwise direction two showed positive deviations and two showed negative deviations.

In each of the slider pitches, the horizontal deviation was positive. The slider seemed to be a combination of the fast ball and the curve ball. The slider had a higher velocity than the regular curve ball although similar forces in the vertical and horizontal planes acted on both types of pitches.

Discussion of Findings

If the velocity and the rate of rotation are disregarded, the direction and the amount of deviation may be considered to have been determined by two factors, namely, the direction of rotation of the ball and the vertical angle. With the exception of the balls in the knuckle-ball category, the balls that rotated in a clockwise direction showed positive vertical deviation and negative horizontal deviation; and the balls that rotated in a counterclockwise

direction showed negative vertical deviation and positive horizontal deviation. In general, as the vertical angle approached 90°, the vertical deviation increased and the horizontal deviation decreased; and as the vertical angle approached 0° or 180°, the vertical deviation decreased and the horizontal deviation increased. For example, in the curve ball category the vertical deviation was negative and the horizontal deviation was positive due to the counterclockwise direction of rotation of the ball. As the vertical angle increased, the vertical deviation generally decreased and the horizontal deviation generally increased as shown in Table 6.

TABLE 6.—A COMPARISON OF THE VERTICAL ANGLES WITH THE VERTICAL AND HORIZONTAL DEVIATIONS FOR PITCHES IN THE CURVE BALL CATEGORY

Vertical Angles (degrees)	Vertical Deviations (ft.)	Horizontal Deviations (ft.)
120	−1.6	.1
130	−1.8	.2
135	−1.6	.9
140	.0	.5
160	.0	.5
160	.0	1.2

In general, the forces that in one plane (i.e., vertical or horizontal) acted upon the ball were inversely related to the forces that in the other plane (i.e., horizontal or vertical) acted upon the ball. The changes in the vertical angle and in the horizontal angle of the pitch resulted in changes in the direction and in the magnitude of the vertical deviation and of the horizontal deviation of the pitch. The changes in the vertical angle and in the horizontal angle during the pitch probably resulted from precession.

The direction of deviation of the knuckle ball was apparently not related to the direction of rotation of the ball. It seems likely that, with such low speeds of rotation, the orientation of the seams relative to the axis of rotation and relative to the air flow during the course of each pitch may have determined the direction of the deviation. The position of the seams of the ball in relation to the axis of rotation of the ball and in relation to the air-flow past the ball may be the reason for the unpredictability of the trajectory of the knuckle ball. This phenomenon may affect the deviation of all types of pitches.

The fast balls consistently displayed positive vertical deviation. The action of the forces that resulted from the clockwise rotation of the pitch apparently caused the fast ball to appear to the batter to hop or rise as it approached the plate. During its flight the ball fell toward the ground, but it did not fall as rapidly as the batter expected it to fall; and thus was created, on the part of the batter, the illusion that the ball rose or hopped during the latter part of its flight.

The sliders, the curve balls, and the sidearm-curve balls consistently had positive horizontal deviations. The sliders (with a velocity of 98fps) had a higher velocity than the curve balls and the sidearm-curve balls for which the velocity was 91fps. The rates of rotation were similar for these three types of pitches. The vertical deviations and the horizontal deviations were also similar for sliders, curve balls, and sidearm-curve balls.

The type of change-up pitches made by the pitchers varied. For the change-up pitches, most of the pitchers threw slow curve balls. The velocities of the change-up pitches were directly related to the velocities of the other types of pitches made by the same pitcher.

Conclusions

1. Each pitch in each category showed either vertical deviation or horizontal deviation or both. None of the pitches followed the course that would be followed by a freely falling body.

2. Within the ranges of the measured rates of rotation and velocities of the pitched balls in this study, the amount of deviation of a pitched ball is determined primarily by the vertical angle.

3. In general, the forces that in one plane (i.e., vertical or horizontal) act upon the ball are inversely related to the forces that in the other plane (i.e., horizontal or vertical) act upon the ball.

4. The relation of air-flow to the seams of the ball affects the magnitude of the deviation of the pitch.

References

1. "Baseball's Curve Balls: Are They Optical Illusions?" *Life* **11**: 83-89; Sept. 15, 1941.
2. Brown, S. Leroy. "Bernoulli's Principle and Its Application to Explain the Curving of a Baseball." *Popular Science Monthly* **83**: 199-203; Aug. 1913.
3. Bunn, John W. *Scientific Principles of Coaching.* New York: Prentice-Hall, 1955. 1955.
4. "Camera and Science Settle the Old Rhubarb About Baseball's Curve Ball.' *Life* **35**: 104-07; July 27, 1953.
5. Coombs, John W. *Baseball.* New York: Prentice-Hall, 1951.
6. "Curve is a Curve." *New Yorker* **25**: 21-22; Oct. 8, 1949.
7. Drury, Joseph F., Jr. "The Hell It Don't Curve." *American Mercury* **76**: 101-06; May 1953.
8. "Fast Action Camera Shows How Hubbell Throws a Curve." *Life* **9**: 6-8; July 22, 1940.
9. Kenny, James D. *A Study of Relative Speeds of Different Types of Pitched Balls.* Master's thesis. Iowa City: State University of Iowa, 1938.
10. "Letters to the Editor." *New Yorker* **17**: 69; June 7, 1941.
11. Madden, R. W. "Our Skeptical Correspondents; Throwing a Curve." *New Yorker* **17**: 61-62; May 24, 1941.
12. McCloy, C. H. *Mechanical Analysis of Activities.* Unpublished course outline. Iowa City: State University of Iowa.

13. MEANY, TOM. "Anyone for Knuckleballs?" *Colliers* **133**: 28; Oct. 2, 1953.
14. ————. "Turley's Fast Ball." *Colliers* **133**: 42-43; June 25, 1954.
15. "The Pitching Repertoire." *Encyclopedia Britannica* **14**: 166B.
16. "Robin Roberts—Control Master." *Look* **17**: 124-27; June 17, 1953.
17. SMITH, STANLEY E. "Watch Those Curves." *Argosy* **342**: 40; Mar. 1956.
18. SPAHN, WARREN, with FAY, BILL. "You Can Fool Some of the Hitters All of the Time." *Colliers* **136**: 62; July 22, 1955.
19. SPALDING, A. G. *Baseball, America's National Game.* New York: American Sports Publishing Co., 1911.
20. VERWIEBE, FRANK L. "Does a Baseball Curve?" *American Journal of Physics* **10**: 119-20; Apr. 1942.
21. "Visual Proof that a Baseball Curves." *Look* **13**: 74-77; July 19, 1949.
22. VOGEL, O. H. *Ins and Outs of Baseball.* St. Louis: C. V. Mosby Co., 1952.

(Submitted 6/27/58)

Notes and Comments

NOTES

On the Use of a Rubber Yardstick in Evaluating a Difference

FRANKLIN M. HENRY
University of California
Berkeley, California

USE OF any objective standard of reference implies that the standard in question be stable and unchanging, rather than elastic. For the purpose of deciding whether an observed difference between two sets of data should be interpreted as simply a reflection of sampling error (null hypothesis) or a "real effect" (alternative hypothesis), statisticians have devised the t-test, the F-test, the Chi-squared test, etc. Any elementary statistical textbook explains the underlying philosophy of these tests and points out that the established critical magnitude of t, F, and χ^2 that divides statistical significance from nonsignificance is that value which yields a 5 percent probability in support of the null hypothesis. (The use of a 1 percent probability, characterized as "highly significant," is also accepted by established practice.) The more sophisticated textbooks consider the consequences of employing a higher (or lower) critical value, and insist that the decision to use some other than the conventional critical level as a basis of decision must be made and justified before the data are examined. The statistical test would otherwise be meaningless.

Unfortunately, authors of RESEARCH QUARTERLY articles sometimes forget the purpose of the tests of significance. Inspection of almost any issue of the QUARTERLY will reveal examples. For instance, on page 44 of the 1958 volume may be seen the following conclusions: "The findings of the study were not found to be statistically significant. . . . Both groups made gains in learning." But these cannot be gains in learning if they are simply fluctuations reflecting sampling error, as established by the first statement. Moreover, the data presented on page 42 have not been analyzed to determine if the gains from learning are statistically significant.

Another example of a conclusion that is at variance with the statistical evalution may be seen on page 114: "While the difference between the two methods was not statistically significant, learning according to the whole method generally required fewer trials." Here again, the two interpretations are mutually exclusive on logical grounds.

On pages 366-67, one reads: "The mean score was significantly higher (+1% level of confidence) . . . The group had a higher degree of motor ability (10% level of confidence) . . . Higher than the group who had pure left dominance (20% level of confidence)." In this example, a floating statistical criterion has been used; the term "level" is a misnomer, and the conclusions drawn are inconsistent with established bases of interpretation.

The issue that has been raised is not concerned with esoteric technicalities. It is a straightforward question concerning the basic purpose of a statistical analysis of a difference and the inescapable consequence of the result of such analysis if a high level of scholarship is to prevail in our research publications. Maintenance of this standard would seem to be the responsibility of reviewing editors as well as authors.

(Submitted 11/6/58)

The Meaning of a Nonsignificant t

A. T. SLATER-HAMMEL
Indiana University

IN USING the t test to evaluate the significance of a difference between two treatment means, the logical basis for rejection of a null hypothesis is to be found in the assessment of probabilities. Should the probability (p) of a particular t be equal to or less than some small preselected value, the researcher would reject the possibility that his difference in treatment means could be attributed to chance variations; he would regard his null hypothesis as being discredited because its truth would assign a small p to his observations. In effect, the researcher rejects a null hypothesis on the principle that what is unlikely to happen by pure chance is unlikely to happen to him.

Since there is no value of p at which it can be said with certainty that differences between means do exist, the rejection of a null hypothesis always involves some risk. However, by selecting his value of p, the researcher can determine the degree of certainty that is to be attached to his statement.

But suppose t falls short of some specified value. Does the nonsignificant t mean that the null hypothesis is true, as many researchers would appear to conclude?

Consider, for example, a recent report in the RESEARCH QUARTERLY. In this investigation, concerned with the problem of lateral dominance in teaching bowling, four left-dominant subjects (experimental group) learned to bowl with the left hand, and 16 left-dominant subjects (control group) learned to bowl with the right hand. Comparisons between groups included measures of general motor ability, bowling averages, and the number of games bowled. The results for the latter measure revealed that the experimental group bowled an average of 11.50 games; the control group, 7.88 games. Since the t for this comparison was only 1.56, falling short of the value 2.88 required for significance at the 1 percent level of confidence, the researcher concluded that "it can be assumed that there was no real difference in the amount of practice between the two groups." But does this form of conclusion necessarily follow? It is the suggestion of this note that the answer is definitely no.

The reason for this answer is that a null hypothesis of zero difference between means is simply an assumption, or argument, set up in the hope that it may be disproved. In effect, the researcher puts the following form of question to his data: On the assumption that the difference between means is zero, how frequently could sampling errors be expected to produce a difference as large as or larger than that found in my observations? Should his test of difference, under this hypothesis, assign a small p to his observations, he would feel justified in concluding that the hypothesis is likely to be false. On the other hand, should p be fairly large he would have no basis for rejecting his hypothesis, and he would conclude that it was tenable. This conclusion, however, would not constitute proof of the hypothesis. It would only signify that the data offer insufficient evidence against the hypothesis, and it would not rule out the possibility that other hypotheses might also be tenable. Hypotheses specifying that the difference between means took various negative or positive values might also be regarded as tenable. Thus, a nonsignificant t is merely a negative result, justifying no more than a verdict of "not proven."

Furthermore, for the class of problems involved in the assessment of differences between treatment means, it is a reasonable assumption that different treatments, if they have any effect upon the experimental material, will not be exactly equal in their effectiveness. For such comparisons a null hypothesis of zero difference will never be true, and a nonsignificant t can only mean that the experimental errors were too great for the effect to be demonstrated.

But in spite of these limitations, it is still possible to draw some useful conclusions from a nonsignificant result. The researcher can always construct a confidence interval which will establish the limits within which hyoptheses about the difference between

means may be regarded as tenable. Then, in place of the unwarranted inference that "there is no real difference between means," the researcher may substitute the valid inference that "the real difference between means will fall inside such and such limits." Although this procedure will entail some risk, as does any statistical inference, the researcher can always select the degree of certainty to be attached to his statement.

Should the confidence limits be small, it might then be concluded that no serious error will be committed in assuming that the difference between means is zero; that is, the probable magnitude of the difference is too small to have practical importance. If, on the other hand, the limits include a fairly wide range of possible values, the only reasonable conclusion is that the data are too inaccurate to show whether there is a difference of any practical importance.

Applying this form of statistical evaluation to the report on lateral dominance in teaching bowling. the confidence interval at the 5 percent level for difference in the number of games bowled would range between —1 and 8 games, at the 1 percent level, between —3 and 10 games. Since these ranges are so broad, the conclusion of "no real difference in the amount of practice between the two groups" would hardly seem justified. In short, the data are simply too inaccurate for any reasonable conclusion about the practical importance of the difference in the number of games bowled.

More important, it is difficult to appreciate the structure of a learning experiment in which differences in the number of games bowled, or the amount of practice, must be statistically assessed. This difficulty becomes acute when one tries to imagine the sort of population from which the researcher, presumably, has drawn two random samples. Logically, the application of a t test to assess such differences would seem justified only if the researcher were concerned with random samples from some population of the possible number of games bowled. But this, obviously, was not the case.

For the actual conditions under which data were collected, it would seem most reasonable for the researcher to have simply admitted the existence of a difference in the amount of practice between groups. Since the group which bowled the greater number of games also had the lower bowling average, this admission need not completely vitiate any conclusion relative to the comparison between bowling performance and the hand used.

However, a significant difference in bowling performance between groups, as reported in this investigation, does not necessarily justify the conclusion that "it would appear that the preferred hand rather than the dominant hand should be used as the bowling hand." This would be a reasonable conclusion only if the groups had been given a fair chance of being initially equivalent, and this does not appear to have been the case. After all, the experimental group consisted of only four subjects, from a pool of 20 left-dominant subjects, who could be "persuaded" to use their left hand rather than their preferred right hand; and the control group was formed from the remaining subjects who were too obstinate to be persuaded from using their preferred hand. (Yes, all "left-dominant" subjects preferred to bowl with their right hand.)

(Submitted 8/18/58)

COMMENT

Comments on the article by Lois Ellfeldt and Eleanor Metheny, "Movement and Meaning: Development of a General Theory," in the October 1958 RESEARCH QUARTERLY.

The new theory of "movement and meaning," though interesting, has several weaknesses. A basic fallacy occurs in the concept of kinesthetic feedback as presented in the statement that "the kinescept provides a sensory record . . . even while it is controlling or guiding the response" (p. 269). The kinesthetic input, or kinescept, is a *result* of the muscular action causing the movement and of the movement. Thus, kinesthesis is an inherent error-sensing mechanism. But the assumption that kinesthesis senses errors in the output in time to alter the action which produces it is highly questionable in the case of fast, skilled movements.

As a simple example, the skilled typist may sense that a wrong key was struck and stop to check visually. Very rarely does the typist block the stroke in time to prevent the error or in time to prevent additional letters appearing. Whether recognition of the wrong key being struck results from a discrepancy between the material being transscribed and the "kinesymbol" or a discrepancy between the "kinesymbol" leading to and the "kinescept" leading from the stroke is a moot question. The important point is that the error initiated tends strongly to occur and that feedback rarely occurs in time to prevent the error. Error prevention, when it does occur, is in terms of a temporary blockage of production. In sport, dance, or piano playing where the situation requires continued production, the error may be noted, but the general process proceeds.

Kinesthesis may be error-sensing, but not error-correcting, in fast movements because of an inherent characteristic of the neuromuscular system. In fast, skilled movements, and especially those involving large segments of the body, the segment outruns the impressed force of the muscle, or muscles, initiating the movement. Thus, control is inherent in the combination of muscular forces initiating the stroke. Even though error is sensed as the movement is being executed, the error results unless the stroke is checked by an antagonist. Then it must be returned and restarted. Compensation for errors sensed through kinesthesis can sometimes be made in subsequent strokes of serial movement patterns. The result may be partially saved. And, of course, errors can be corrected in subsequent executions of the skill. In general, though, kinesthetic input is like a follow-through—it may indicate what went wrong but not in time to prevent it.

A second major weakness of the new theory is the suppression or disregard of the vital part other sensory inputs have in shaping skilled motor performance. The definitions (p. 268) establish a hierachy from action to concepts based on kinesthesis, but kinesthesis without touch, pressure, and vision would be a relatively sterile sense. The quality, meaning, and effectiveness of motor performance depends on visible effects. Vision generally provides the basic feedback concerning the external effect of our actions. Vision also provides a basis for comprehending the action of others. Many "kinesymbols" originate in visual perception—seeing and trying are normal. Visual analysis of motor performance by skilled observers (coaches) often locates the source of motor errors and leads to better performance—and probably better "kinesymbols"—through discourse and demonstration.

Kinesthesis, touch, and pressure inputs accompany the production of action and precede the visible effect. A performer can predict the visible result from these inputs —once he learns to interpret them. He may "feel" the difference between two performances, but he cannot perceive good movement kinesthetically until he produces it. He cannot predict the outcome of performance from the "feel" during production until he determines what cues from kinesthesis, touch, and pressure correlate with good performance. Otherwise, old errors feel good. Obviously, vision is necessary to keep

abreast of play in competition with opponents. But an individual executing a thoroughly practiced routine in gymnastics or figure skating, where he has only himself to control, might conceivably operate on the basis of cues from kinesthesis, touch, and pressure. He can, but partial or complete deprivation of vision produces considerable decrement in performance (1). This suggests that kinesthesis, even with touch and pressure, is not a sufficient basis for controlling previously well "kinesymbolized" skills.

A third weakness may be simply overstatement for emphasis. Statements that "a kinestruct can never be described in detail," that "the 'feel of a movement' can never be described in words," and that "conceptualizations of kinesthetic perception cannot be expressed in the symbols of any other sensory conceptualizations" are highly questionable. "He raised his arm" is concise and presumably sufficiently descriptive for the occasion. Describing any movement "in detail" is infinitely complicated and probably useless. Dynamic forms can be described, although precise description may confuse the unitiated. Incidentally, human movement is compounded from discrete muscle impulses rather than "continuous changes in tension in every muscle fiber of the body." The "feel of movement" can be shared. Prefacing statements with "it should feel as though" directs attention to kinesthetic, touch, and pressure sensations and, by directing a person to try for this feel, often bears fruit. The "kinesymbol," "kinescept," and "kinestruct" may be unexpressed operationally, but this does not mean that they are unexpressable or uncommunicable. Speaking is a motor response and thinking is done in terms of subovert movements. We translate words into action and action into words. We verbalize visual concepts, visualize verbal concepts, and express ideas with movements. And if a dancer can communicate concepts without discourse, discourse might communicate concepts without large, somatic movements—granting that both might miss the point in translation and both translations might result in jargon.

Finally, "recognizing that the kinestruct and its kinescept are both kinesymbols" and that these "cannot be expressed in the symbols of any other sensory conceptualization" makes all action and errors kinesymbolic and untouchable. Automatic feedback control of kinestructs by kinescepts makes action reflex and errors immutable since the movement is a motor image of the symbolic pattern. Thus, in trying to summarize the authors' position, but with no intention of distorting their meaning by injudicious quotation, we seem to find that movement has an untouchable origin and an immutable nature which presents physical education with an unsolvable problem and which makes any claims of teaching or training people in motor skills presumably fraudulent. The general theory (p. 272) might be rephrased to state that human movement is meaningful action based on the meaning of movement. This does not summarize the extensive findings concerning movement, perception, and their relations, but it at least leaves the door open for learning.—*Alfred W. Hubbard, University of Illinois, Urbana, Illinois.*

References

1. Graybiel, Ashton, *et al.* "Russian Studies of Vision in Relation to Physical Activity and Sports." *Research Quarterly* **26**: 480-85; Dec. 1955.

(Submitted 11/20/58)

Research Abstracts

Prepared by the Research Abstracts Committee of the National Council of the Research Section, D. B. VAN DALEN, Chairman

28. BLAKESLEE, ALTON L. "Where Do We Stand on Cancer, Heart Disease, and Mental Illness?" *Today's Health,* **36:** 16-20, 53; Aug. 1958.

Twenty years ago only one person in seven was saved from cancer while today it is one in three. Great progress has been made in the development of drugs, but one should not expect a single miracle anticancer drug. Cancer appears to be a group of diseases; therefore, a single drug may not be effective. Vaccines are being developed by finding specific chemical differences between cancer and healthy cells and by using those chemicals to stimulate the human body to build up defenses against cancer cells. Reliable and simple tests should be developed to detect cancer early so prompt chemical, surgical, or radiation treatment can be started far earlier.

The common cause of heart attacks is atherosclerosis, a fattening or narrowing of the inner walls of the coronary arteries. Experts are not in agreement as to the cause of such attacks. Some state they are due to too much fatty food in the diet, others place the blame on cholesterol, while others state the attacks may be due to a lack of exercise, tension, or smoking. Anti-coagulants are reducing the danger that clots will form to produce attacks in heart patients. Blood clots blocking heart arteries are being dissolved by certain enzymes. Drugs have been developed to lower blood pressure. Surgeons now routinely repair damaged heart valves. Amazing results have also been obtained through heart massage and electrical shocks injected directly into the heart. Heart-lung machines have also proved to be invaluable to the heart surgeon.

The new concept in the treatment of the mentally ill is the growing realization that their well-being is a responsibility to be shared by the hospital and the community. Many feel mental illness can result from a defect in the way a person's body handles normal chemicals. Drugs used in the treatment of mental illness are being improved, with much of the research aimed at learning exactly how tranquilizers and energizers work. The brain is being explored electrically with fascinating results. Occupational therapy in groups is another excellent means of aiding the mentally ill.—*J. Grove Wolf.*

29. CLAWSON, MARION. "Recreation Land Resources for the Year 2000." *Recreation* **52:** 12; Jan. 1959.

In a paper prepared for the 40th National Recreation Congress, the author predicts that by the year 2000 the United States will have increased its population to about 310,000,000; and, in order to meet the recreation needs, our user-oriented areas will have to be four times greater; the intermediate areas 16 times greater; and the resource-based areas 40 times greater than they are now. Personnel, acreage, and policy issues must receive attention now in order to be adequate for the future.—*Elizabeth M. Prange.*

30. COHEN, WALTER; HERSHKOWITZ, AARON; and CHODACK, MAJORIE. "Size Judgment at Different Distances as a Function of Age Level." *Child Development* **29:** 4; Dec. 1958.

Forty-two subjects were randomly selected from various classes of the Barker, N. Y., school system. The experimental group of 42 boys and girls was divided as follows: 12 subjects in each of the 7-, 12-, and 17-year-old groups and six subjects in the 5-year-old group. Judgments between a comparison object placed 8 meters from the subjects and a standard placed either 2 meters or 6 meters from the subjects were obtained from all

of the subjects. The 17-year-olds were significantly more accurate in size estimation than the younger groups (5-, 7-, and 12-year-olds) when the standard was at 2 meters from the subjects but not when it was at 6 meters. The 17- and 12-year-olds had a lower I.U. (interval of uncertainty) than the 5-year-old at both distances. These results suggest that the distance between test objects is an important variable in understanding the contradictory data obtained in developmental studies of size constancy. One of the major implications of the study is that size constancy should not be treated as a unitary phenomenon.—*D. B. Van Dalen.*

31. DeKruif, Paul. "A Million Hidden Diabetics." *Today's Health* **36:** 22-23, 68; Nov. 1958.

Diabetes remains the seventh leading cause of death by disease. Over a million persons in the United States have the disease but do not know it. In normal people, starchy and sugary foods, which are the chief sources of energy, are burned immediately or stored in the body. When the body can't handle these carbohydrates, sugar piles up in the blood and spills over in the urine. This is diabetes. Inheritance plays a significant part in the development of diabetes. A potential diabetic should avoid an excessive intake of sweets and foods high in carbohydrate. He should avoid overeating, as obesity has been identified as the spark that sets off the diabetes explosion in patients. Diabetes found early can be checked by use of insulin, new oral remedies, and a somewhat restricted though adequate diet. Authentic information on diabetes can be secured from the American Diabetes Association, 1 E. 45th St. New York.—*J. Grove Wolf.*

32. Duel, Henry J. "Effect of Periodical Self-Evaluation on Student Achievement." *Journal of Educational Psychology* **49:** 4; Aug. 1958.

This study was undertaken to determine if self-evaluation is of value in improving student achievement. The subjects were Air Force enlisted students in an electronics communications course from two technical schools at Scott Air Force Base. The measuring instrument (set) was simply a scale upon which each student was asked to rate himself. Subjects in the two schools evaluated their skills periodically, under careful control of the experimenter in one school and as part of the "normal" class activities in the second. Their achievement tests were compared with control groups, and the results favored the self-evaluation groups in both schools. Students given the opportunity to evaluate themselves can achieve to a greater degree than students not receiving the same opportunity.—*D. B. Van Dalen.*

33. Gerall, Arnold A., and Green, Rueeel F. "Effect of Torque Changes upon a Two-Hand Coordination Task." *Perceptual and Motor Skills* **8:** 217; Dec. 1958.

This study was initiated to determine whether performance on a tracking task is influenced by loading demands during earlier training periods. The task consisted of tracking a slowly moving dot on the face of a cathode ray oscilloscope. The subjects governed movement of a target follower, a circle, by means of two 2-in. control cranks. An electromagnetic brake determined the amount of force necessary to produce rotation. Twenty-six subjects were selected from a military population and randomly assigned to one of two groups. One group of 13 subjects was required to work against 2 lb. of coulomb friction during original practice and 14 lb. of transfer. A second group had the frictional loadings reversed. The results shows a relatively large and persistent performance decrement with a change from light to heavy loadings. With the change from heavy to light loading, decrement was smaller and transient.—*D. B. Van Dalen.*

34. Hase, Gerald J. "Where Are the Accidents Happening?" *Safety Education* **38:** 40; Oct. 1958.

Results of a survey on accidents in 92 New York state elementary schools showed that

in the physical education class instruction program the intermediate grades had the highest accident rate per 100,000 exposures. Wrestling, specifically, had the highest rate; and, in the general categories, games had a rate of 4.7 accidents per 100,000 exposures followed in order by stunts and tumbling, gymnasium apparatus, rhythms and dance, body mechanics, and swimming. Extra-mural activities had a higher accident rate than class instruction or intra-mural activities.—*Elizabeth M. Prange.*

35. Hughes, E. L. "Standardized Weight Training Performance Test." *Athletic Journal* **39**: 34, 36, 55-58; Oct. 1958.

No standardized weight performance test has been published in the literature; therefore, the University of Washington decided to develop its own test. Weight exercises used included sit-ups, pull-overs, curls, military press, rowing motion, bench press, squats, stiff deg-dead weight lift, chins, snatch, and the clean and jerk exercise. Standards of performance were based on a percentage of body weight and normative tables developed based on the number of repetitions for each particular exercise.—*J. Grove Wolf.*

36. Liddle, Gordon. "Overlap among Desirable and Undesirable Characteristics in Gifted Children." *Journal of Educational Psychology* **49**: 4; Aug. 1958.

This study was made to examine the overlapping of talents and maladjustments in a groups of 1015 public school children in late childhood and early adolescence in a mid-western city of 45,000. The following characteristics were measured: aggressive maladjustment, social leadership ability, artistic talent, and intellectual ability. Various tests and the "Behavior Descriptive Chart" were used as measuring devices. The top 10% in each group listed above was withdrawn and examined. It was found that children who were highly gifted in one of the three talent areas were quite likely to be talented in other areas, and quite unlikely to be seen as highly maladjusted by their teachers and classmates.—*D. B. Van Dalen.*

37. Jones, Mary Cover, and Mussew, Paul Henry. "Self-Conceptions, Motivations, and Interpersonal Attitudes of Early and Late-Maturing Girls." *Child Development* **29**: 4; Dec. 1958.

The relationship between maturational status and TAT scores was determined for a group of physically accelerated as contrasted with a group of slow developing girls from a normal classroom sample. The experimental group of 34 girls consisted of 16 accelerated girls and 18 retarded girls. The girls were analyzed according to a scoring scheme involving 20 needs, press, and descriptive categories. An analysis of the data found few striking differences between the two groups of girls in attitudes and motivations. However, early maturing girls had scores that indicated more favorable self-concepts. The findings were interpreted to indicate that late maturing adolescents of both sexes (boys from a previous study) are characterized by less adequate self-concepts, slightly poorer parent-child relationships, and some tendency for stronger dependency needs.—*D. B. Van Dalen.*

37. Lipsitt, Lewis P. "A Self-Concept Scale for Children and its Relationship to the Children's Form of the Manifest Anxiety Scale." *Child Development* **29**: 4; Dec. 1958.

The children's form of the manifest anxiety scale and self-concept and ideal-self scales were administered twice to approximately 300 fourth, fifth, and six graders at a two-week interval. A discrepancy score, or measure of self-disparagement, was obtained by subtracting S's self-concept ratings from their ideal-self rating. It was found that the self-concept measure taken by itself was more reliable than the discrepancy measure, and that the self-concept measure was more highly related to CMAS (children's manifest anxiety scale) score than was the discrepancy score. Significant correlations were obtained for all grades and sex combinations between CMAS and self-concept scores, with high

anxious S's producing low self-concept ratings. The two-week reliability of the CMAS was essentially the same as that for the original population.—*D. B. Van Dalen.*

38. MATHEWS, D. K.; SHAW, VIRGINIA; and RISSER, PHILIP. "The Moses Lake Project." *Journal of Health, Physical Education, and Recreation* **29**: 18; April 1958.

In an effort to determine fitness in relation to psychological stability, health, body mechanics, and physical anthropometry, the state of Washington started a fitness program at Moses Lake Junior High School in Moses Lake, Washington. Using a group approach, the Rogers' PFI and the Kraus-Weber tests of minimal muscular fitness were administered to some 1,000 boys and girls. The 40 low scoring children were assigned to a special program for the sub-fit child.—*Elizabeth M. Prange.*

39. MICKELSEN, OLAF. "Age Changes in Body Composition." *Public Health Reports* **73**: 295-301; April 1958.

Methods of determining body composition and the factors influencing the composition are discussed with reference to several studies made in those areas. Available data seems to indicate an increase in body fat up to about age 50 with women gaining more than men. Physical activity is associated with a marked reduction of body fat. Since women exceed in body fat but still outlive men, it is suggested that the secret to a long life for a body that tends toward obesity may be found by studying older women.—*Elizabeth M. Prange.*

40. MILLER, K. M., and BIGGS, J. B. "Attitude Change through Undirected Group Discussion." *Journal of Educational Psychology* **49**: 4; Aug. 1958.

Attitude changes are subject to many variables. This study was concerned with the effectiveness of free group discussion about racial groups when the discussion groups are sociometrically structured. The subjects consisted of two small groups of boys from two senior secondary school classes. The groups selected were (1) psychogroup considered high, (2) sociogroup low in cohesion. A control group was also used in the study. Several findings were definite while others merit further investigation. Free, undirected discussion about racial groups by two types of small groups, selected on a sociometric basis, resulted in a significant change of attitude irrespective of the type of group. No differences between quantitative changes of friendly and neutral S's were discovered.—*D. B. Van Dalen.*

41. MITCHELL, JERE H., *et al.* "The Physiological Meaning of the Maximal Oxygen Intake Test." *Journal of Clinical Investigation* **37**: 538-47; April 1958.

A maximal oxygen intake test was given to normal subjects using a treadmill and using the point when the oxygen intake curve ceased to rise (when plotted against the load of work) as the maximum intake. The oxygen increased about 9.5 times, the cardiac output 4.3 times, the pulse rate and stroke volume 2 times. The ability for the A-V oxygen difference to widen seemed to be the important factor of maximum intake.—*Ernest D. Michael.*

42. OLSON, H. W. "The Effect of a Supervised Exercise Program on the Blood Cholesterol of Middle-Aged Men." *The Physical Educator* **15**: 135-137; Dec. 1958.

A control group of 15 subjects with a mean age of 40.53 years followed a sedentary life while another group of 16 subjects averaging 43.06 years were given an individually prescribed group of exercises over a period of three months. The former group included two subjects with high serum cholesterol while the latter had three such subjects.

No appreciable differences were found in blood serum cholesterol between the mean, initial, and final values in either group. However, individually, there were quite large changes in both the exercise and control groups. Some subjects increased, some decreased.

In the high cholesterol subjects all three of the exercise subjects showed a decrease in blood cholesterol. Of the two controls, one showed an appreciable drop in blood cholesterol and the other did not change. In either group, subjects who showed an increase in blood serum cholesterol were generally found to show an increase in body weight.—*J. Stuart Wickens.*

43. Sarason, I. G. "Effects on Verbal Learning of Anxiety, Reassurance, and Meaningfulness of Material." *Journal of Experimental Psychology* **56**: 6; Dec. 1958.

The anxiety of 76 subjects was assessed by a Test Anxiety Scale (TAS) developed by the writer and by Bendig's short form of the Taylor Manifest Anxiety Scale (MAS). The subjects with TAS scores of six and over were assigned to high TAS groups; those with scores below six to low TAS groups. Twelve randomly selected subjects were dropped in order to have eight subjects in each treatment group and thereby make it possible to carry out analysis of variance involving repeated measurements. Two of the high TAS groups received standard instructions and two of the high TAS groups received standard instructions plus reassuring instructions. One of each of these two sub-groups was given lists of words having high meaningfulness; the other, lists having low meaningfulness as determined by Noble's *m* values. The low TAS groups were subdivided and treated in the same way.

It was found that reassurance instructions were relatively facilitative in the case of the high TAS subjects, but relatively detrimental in the case of the low TAS subjects. The performance of the low TAS subjects appeared to be facilitated by standard instructions, whereas the opposite seemed to be true of the high TAS subjects. When data were analyzed in terms of the Taylor MAS scores, no effects involving anxiety were found to be significant—*Edna Willis.*

44. Sarason, Seymond B., *et al.* "Classroom Observation of High and Low Anxious Children." *Child Development* **29**: 2; June 1958.

Thirty-two matched pairs of high anxious (HA) and low anxious (LA) children comprised the experimental study. The Rorschach, parental interviews, and direct observations in the classroom were employed. Satisfactory reliability of observations was established. The HA boys differed from LA boys in that they did not appear as academically adequate and showed less task orientation and greater insecurity. HA girls differed from LA girls in that they seemed to have a stronger need for achievement, showed less unintelligible behavior, and contained fewer distractible individuals.—*D. B. Van Dalen.*

45. Schultz, Richard E. "Patterns of Personal Problems of Adolescent Girls." *Journal of Educational Psychology* **49**: 1; Feb. 1958.

A cluster analysis was performed on 156 selected items from the Billet-Starr Youth Problems Inventory, Senior Level, based on the responses of 500 adolescent girls. Three clusters were extracted and designated: Cluster I, general personal anxiety and insecurity; Cluster II, tension concerning relations with others; and Cluster III, difficulties in getting along with parents.—*D. B. Van Dalen.*

46. Smith, Olin W., and Gruber, Howard. "Perception of Depth in Photographs." *Perceptional and Motor Skills* **8**: 307; Dec. 1958.

This study had two tasks: (1) Determine the empirical function describing the relationship between the distance of viewing a photograph of a 3-dimensional scene and the depth apparent in it. (2) Attempt to explain why O usually perceives a single specific space when more than one space can be geometrically demonstrated to be in correspondence with a fixed projection. A photomural was viewed monocularly with a field of view restricted to the photograph when observed from six different distances (1.0 to 2.8 meters) by 20 O's (O's were 20 Cornell students with at least 20/20 vision), who made 5 judgments at each distance. The apparent depth in the portrayed scene at each distance

of observation was compared by means of ratio judgments with the apparent depth in the scene which had been photographed. Judgments of depth varied as a function of the distance from which the photograph was viewed, but the height and width of the scene remained constant. A determinant for the constancy of apparent height and width was proposed.—*D. B. Van Dalen.*

47. Smith, Patrica Ann. "Emotional Variables and Human Motion." *Perceptual and Motor Skills* **8**: 3; Sept. 1958.

The performance on a panel control task of 35 subjects was studied to learn the effects of emotional variables on skilled human motion. The subjects were divided into five experimental groups in terms of psychometrically defined levels of anxiety and rigidity. Electronic techniques were used to measure the duration of the travel and manipulation components of motion. Emotional variables influenced the learning of travel movements. However, significant differences among groups in terms of level of performance were found only for the manipulation component, and these differences appeared to be related primarily to the measure defined as rigidity. No differences were noted on either component of motion at a skilled level.—*D. B. Van Dalen.*

48. Shepard, A. H., and Abbey, D. S. "Manifest Anxiety and Performance on a Complex Perceptual-Motor Task." *Perceptual and Motor Skills* **8**: 327; Dec. 1958.

This study investigated the relationship between manifest anxiety and performance on a complex perceptual-motor task. A group of 483 undergraduates was tested for emotional stability. Two groups differing in manifest anxiety were determined and from each group, 28 students (10 male and 18 female) were selected. Manifest anxiety in relation to level of performance was investigated by using the Tronto Complex Coordinator. The subjects practiced 24 1-minute trials which were divided into 3 blocks of 8 minutes separated by 5-second intervals. It was found that 28 non-anxious subjects were superior in terms of both a higher number of matches and a lower error-match ratio. Males showed superior performance on both measures.—*D. B. Van Dalen.*

49. Stern, Samuel. "The Ballistocardiogram in Overweight Young Adults." *Circulation* **17**: 87-89; Jan. 1958.

A group of young adults found to be normal on electrocardiogram, pulse rates, blood pressure, and x-ray tests were given a ballistocardiogram. It was found that the overweight group had abnormal ballistocardiograms, suggesting also that there was a direct relationship between the severity of the picture and the amount of overweightness.—*Ernest D. Michael.*

50. Tyler, Bonnie B. "Expectancy for Eventual Success as a Factor in Problem Solving Behavior." *Journal of Educational Psychology* **49**: 3; June 1958.

It was hypothesized that an expectancy for eventual success in solving a particular problem could be established which would in turn affect whether the problem would be solved. One group of subjects received encouragement, the second discouragement, the third a combination of the two, and a fourth group served as a control. Encouragement and no comment (control) are both superior to discouragement and to intermittent encouragement and discouragement. More subjects in the low expectancy groups than in the high expectancy group attempt to memorize a solution to the problem in contrast to working out a logical solution. There was an inverse relationship between an expectancy for immediate positive reinforcement and the decision time required for making a response.—*D. B. Van Dalen.*

51. Van Dalen, D. B. "The Role of Fact and Theory in Research." *Physical Educator* **15**: 105-106; Oct. 1958.

This article concerns itself with the interrelationship between theories and facts. They

interact constantly and are inextricably interwoven, with one depending upon the other. Facts without theory or theory without facts lack significance. Facts take their significance from the theories which define, classify, summarize, and predict them. Theories possess significance when they are built upon, clarified, and tested by facts. Consequently, the growth of science is dependent upon the accumulation of pertinent facts and the formulation of new or broader theories.—*J. Stuart Wickens.*

52. Walters, Raymond. "Statistics of Attendance in American Universities and Colleges, 1958." *School and Society* **86**: 2142; Dec. 1958.

The 1958-59 enrollments in American universities and four-year colleges form the largest total in United States academic history. The increase is due primarily to an increase of 4.1% among full-time students. Aiding this 4.1% increase was the high 7.1% increase among freshmen. There is no striking difference in enrollment in accredited universities and colleges on a geographical basis. The most startling drop in enrollment appeared in engineering freshmen. Seven out of ten of a list of 127 institutions which teach engineering reported decreases of varying degrees. There was a 7.6% decrease among full-time engineering students. This current decrease is expected to spark a fresh campaign to attract additional qualified young Americans into the engineering field.—*D. B. Van Dalen.*

53. Wang, Cing Hsi. "The Galvanic Skin Reflex: A Review of Old and Recent Works from a Physiological Point of View." *American Journal of Physical Medicine* **37**: 35; 1958.

This is part two and the conclusion of an excellent summary on methods and studies on the galvanic skin reflex.—*C. Etta Walters.*

54. Wilson, William Cody. Imitation and the Learning of Incidental Cues by Pre-School Children." *Child Development* **29**: 3; Sept. 1958.

The general phenomenon of learning by imitation has two components: one, learning of a new response by matching a model's response, and two, the performance of the response in an appropriate situation in the absence of the model. The second component is the primary focus of this study. The subjects were 26 preschool children from the metropolitan Boston area. The children were from upper middle class families with a median I.Q. of 130. It was hypothesized that the process of learning an imitated response in an appropriate situation in the absence of the model was essentially that of learning of "incidental cues." The children did utilize the model's response as a primary cue for the performing of the same response, but they also learned, in the absence of instructions, a secondary "incidental" cue for the performance of that response.—*D. B. Van Dalen.*

55. Yarbrough, Mary E., and McCurdy, Howard. "A Further Note on Basal Metabolism and Academic Performance." **49**: 1; Feb. 1958.

Evidence contradicts the hypothesis that basal metabolic rate might affect academic performances in college. The correlation between the psychological and intellectual indices used appears to be at or near zero.—*D. B. Van Dalen.*

56. Zaoussis, A. L., and Janes, J. I. P. "The Iliac Apophysis and the Evolution of Curves in Scoliosis." *Journal of Bone and Joint Surgery* **40B**: 442-53; Aug. 1958.

Two hundred and twenty-four patients with scoliosis were studied to determine the radiologic characteristics of the iliac apophysis, its relationship to the growth of the vertebral bodies, and the progression of the curve in scoliosis. The results confirm Risser's observations (1948) that the sensation of spinal growth and curved progression coincides with the completion of growth in the iliac apophysis.—*Wayne D. Van Huss.*

Criteria for Evaluation of Articles Submitted for Publication in the Research Quarterly

General Editorial Policy

1. Material suitable for publication shall include (a) reports of research studies, (b) reviews of research, (c) critical comment on published research, (d) descriptions of new research apparatus or techniques, and (e) abstracts and book reviews. Lists of theses or of bibliographical titles shall not be accepted.

(a) Research studies may utilize any sound method of investigation, e.g., experimental, historical, philosophical, survey. The research shall not have been published previously. Reports based on research for which the magisterial or doctoral degree has been granted shall be judged, for content and form, by the same criteria presented herewith for evaluation of other reports.

(b) Reviews of research shall be critical, comprehensive, and currently significant.

(c) Preliminary reports of research, critical comment on published research, and descriptions of new research apparatus or techniques shall be included in the Notes and Comments section of the RESEARCH QUARTERLY.

(d) Abstracts of research and reviews of scientific and scholarly books shall be included in a special section of the RESEARCH QUARTERLY.

2. Membership of the author in the AAHPER or in any other organization or institution shall not influence acceptance or rejection of a manuscript submitted.

3. Manuscripts of research reports and reviews and of material for the Notes and Comments section shall be reviewed independently by at least three associate editors whose familiarity and competence in the area of the research presented are well established.

Approval of at least two of the associate editors is necessary for acceptance. While the decision of the associate editors is final, provision should be made for reconsideration by the same or other associate editors if the author seems to have basis for appeal.

Material for the Research Abstract and Book Review section shall be reviewed by at least one associate editor.

4. Manuscripts accepted shall be published in order of submission. However, exceptions may be necessary if acceptance of the manuscript is depend-

ent upon revisions required by the associate editors and these revisions are not made promptly.

5. The author shall be notified by the managing editor at the time of acceptance of the approximate date of publication.

6. No changes, additions, or deletions shall be made in any article without approval of the author.

General Statement of Function of Associate Editors

The guiding principles of the associate editor in reviewing and judging manuscripts shall be (1) to maintain a high standard of scholarship in the publication and (2) to encourage and assist research workers in meeting these standards.

The associate editor should be competent in the area of study of the manuscript under consideration—familiar with factual data and with research methods and tools available; sensitive, open-minded, and critical in recognizing new ideas and methods and in evaluating new data. He should be thoroughly objective and impartial in his review, basing judgment solely upon evidence presented, not upon opinion of either associate editor or author. He must recognize and control an almost inevitable human tendency to bias and avoid any hint of censorship in his judgment.

The associate editor should be alert in recognizing intrinsic merit and potentially valuable contribution in a study which may be poor in its presentation or incomplete in its development. A constructive, sympathetic attitude on the part of the associate editor, with positive suggestions to the author for revision or extension of either the report or the research represented, will help to salvage both the accomplished research and the worker. An overly harsh attitude may lose both to the profession.

The associate editor in accepting his appointment should be fully informed and aware of the responsibility involved and the importance of carrying out his assignments not only with good judgment and fairness but also with dispatch.

If an associate editor does not feel adequately qualified to review a particular article, he should return the article to the managing editor immediately, with recommendation that it be referred to another associate editor for review.

If for any reason he is unable to devote the necessary time and effort to do good reviewing promptly, he should take the initiative in resigning as associate editor.

Specific Criteria for Evaluation with Respect to Content

1. The merit of the research article shall be judged in terms of its contribution to knowledge and to the development of theory, either as new information, substantiation, or contradiction of previous findings or application of new or improved research techniques to old problems. It should be recognized that pioneering research may sometimes be concerned with the securing of basic data, which may afford a basis for the later development of theory.

2. The subject matter should be pertinent to health, safety, physical education, and/or recreation, either as applied or basic research.

3. The data presented should warrant a reasonably definite conclusion, i.e., should not be merely a progress report. However, the extent or comprehensiveness of research reported should not be the sole basis for acceptance or rejection; a report of the early stages of a research study may be valuable for the information per se and/or as a guide to other workers.

4. Good scholarship should be evidenced both in the design and execution of the study and in the interpretation of the data presented.

Following is a suggested list of important elements in a well-conducted investigation:

(a) The problem is significant, related to the development or refinement of theoretical principles and/or the testing of hypotheses.

(b) The problem is adequately delimited and the specific purposes of the research are clearly stated.

(c) The methods used are appropriate and adequate to yield data necessary for the solution of the problem.

(d) The methods and procedures used in collecting the data are clearly described with evidence of precision of measurement included, whenever reasonable.

(e) If the problem is such that quantitative analysis of the data is indicated, the statistical methods used are appropriate and adequate.

(f) In an historical study, accuracy, thoroughness, objectivity, and sound logic should be evidenced in the collection, analysis, and synthesis of the data. Emphasis should be on analysis, synthesis, and interpretation, rather than on mere reporting.

(g) A philosophical study, dealing with values or new concepts, should be based on a thorough, logical, cogent, and creative development of rationale. It should be developed in relation to a clearly stated hypothesis. It should not be merely superficial speculation. It should avoid "circular reasoning."

(h) In a survey study, emphasis should be on exposition and analysis, with particular attention to conceptualization and the appropriate selection and adequate sampling of the relevant population. The significance of the findings for diagnosis, policy, or program should be indicated and justified.

(i) In all studies there should be adequate documentation; the data should lead to a definite set of conclusions; the conclusions and interpretations should be based upon and supported by the data presented. Relevance of the findings to theory and to the testing of hypotheses should be indicated. Discussion or theorizing should be clearly identified as such.

(j) For a more extensive guide to evaluation, reference may be made to Carmichael[1] or similar source.

5. Review articles shall be judged on the bases of current significance of the subject matter, comprehensiveness, and the quality of critical analysis and skill in integration shown by the author.

6. The Notes and Comments section provides opportunity for brief reports, not exceeding 500 words, of a preliminary study or of an observation or finding, which might have been incidental to another study or one which the author does not intend to pursue further. This section may also serve as an

---

[1] L. Carmichael, *Manual of Child Psychology*, New York: J. Wiley and Sons, 1946.

open forum for brief, objective, well-considered, critical comment on research previously reported in the RESEARCH QUARTERLY, and for the author's reply to such comment.

Clear, concise descriptions and/or diagrams of new research apparatus or techniques developed by the author, or generally inaccessible, are suitable for inclusion in this section, provided that reasonable demonstration of their value is given. Apparatus and methodological articles are not necessarily limited to 500 words.

7. The Research Abstract and Book Review section is designed for the brief presentation of pertinent research material from other publications, both foreign and domestic, in physical education and related fields of study. Abstracts of degree dissertations may be included here.

Specific Criteria for Evaluation with Respect to Form and Organization

1. Material shall be presented in a clear, logical, orderly, objective way.
2. Manuscripts submitted shall conform in technical details to the form prescribed by the managing editor. A description of technical details, i.e., bibliographical form, tables, illustrations, kind of paper, etc., shall be formulated by the managing editor, with the advice of the Research Quarterly Advisory Committee, and given adequate publicity.
3. The material shall be presented in the briefest form consistent with clarity, accuracy, and completeness. Standard statistical formulas should not be given; commonly used techniques should not be described in detail.
4. Correct use of language, as well as a certain dignity and objectivity in literary style, shall be expected.

Committee on Criteria for Evaluation of Articles Submitted for Publication in the Research Quarterly
Karl Bookwalter
David K. Brace
Esther French
Franklin Henry
Eleanor Metheny
Lawrence Rarick
Pauline Hodgson, *chairman*

March 1959

Check List of AAHPER Research Publications

RESEARCH METHODS APPLIED TO HEALTH, PHYSICAL EDUCATION, AND RECREATION

Techniques and methods of conducting research in these fields. Basic text and reference for courses in research methods. Second edition, 1959. 535 pp. $6.00.

MEASUREMENT AND EVALUATION MATERIALS IN HEALTH, PHYSICAL EDUCATION, AND RECREATION

Prepared by the AAHPER Research Council, this is a companion volume to *Research Methods*. Summary of useful measurement and evaluation instruments to determine program outcomes. 1950. 150 p. $1.00.

EVALUATION STANDARDS AND GUIDE

Revised edition of NCATE standards and guide in health, physical, and recreation education. 1959. 32 p. $1.00.

MASTERS THESES IN HEALTH, PHYSICAL EDUCATION, AND RECREATION

Edited by T. K. Cureton. Over 4000 theses, completely cross-indexed by subjects and areas. Of special value for students, teachers, and research workers. 1952. 296 p. $3.00.

AMERICAN ACADEMY OF PHYSICAL EDUCATION

Papers and reports presented at annual meetings.

Professional Contributions No. 2. 1952. 88 p. $1.50.
Professional Contributions No. 3. 1954. 144 p. $1.50.
Professional Contributions No. 4. 1955. 99 p. $1.50.
Professional Contributions No. 5. 1956. 69 p. $1.50.
Professional Contributions No. 6. 1957-58. 160 p. $2.00.

Order from: AAHPER, 1201 - 16th St., N.W., Washington 6, D. C.